

# Motif Statistics and Spike Correlations in Neuronal Networks

Yu Hu<sup>1</sup>, James Trousdale<sup>2</sup>, Krešimir Josić<sup>2,3</sup>, and Eric Shea-Brown<sup>1,4</sup>

<sup>1</sup>Department of Applied Mathematics, University of Washington, Seattle, WA 98195

<sup>2</sup>Department of Mathematics, University of Houston, Houston, TX 77204-5001

<sup>3</sup>Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001

<sup>4</sup>Program in Neurobiology and Behavior, University of Washington, Seattle, WA 98195

March 8, 2013

## Abstract

Motifs are patterns of subgraphs of complex networks. We studied the impact of such patterns of connectivity on the level of correlated, or synchronized, spiking activity among pairs of cells in a recurrent network model of integrate and fire neurons. For a range of network architectures, we find that the pairwise correlation coefficients, averaged across the network, can be closely approximated using only three statistics of network connectivity. These are the overall network connection probability and the frequencies of two second-order motifs: diverging motifs, in which one cell provides input to two others, and chain motifs, in which two cells are connected via a third intermediary cell. Specifically, the prevalence of diverging and chain motifs tends to increase correlation. Our method is based on linear response theory, which enables us to express spiking statistics using linear algebra, and a resummation technique, which extrapolates from second order motifs to predict the overall effect of coupling on network correlation. Our motif-based results seek to isolate the effect of network architecture perturbatively from a known network state.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Neuron models, cross-spectra, and measures of collective network activity</b>	<b>7</b>
2.1	Networks of integrate-and-fire neurons . . . . .	8
2.2	Measure of network correlation . . . . .	8
2.3	Linear response approximation of cell response covariance . . . . .	9
2.4	Applicability of linear response theory . . . . .	11

<b>3</b>	<b>Graphical structure of neuronal networks</b>	<b>12</b>
3.1	Network motifs and their frequencies . . . . .	12
3.2	Generating graphs with given motif frequency . . . . .	14
<b>4</b>	<b>Impact of second order motifs in networks of excitatory cells</b>	<b>15</b>
4.1	Results: linear dependence between mean correlation and motif frequencies .	15
4.2	First theory: second order truncation for network correlation . . . . .	16
4.3	Improved theory: resumming to approximate higher-order contributions to network correlation . . . . .	18
4.4	Correlations in external input . . . . .	24
<b>5</b>	<b>Impact of second order motifs in networks of excitatory and inhibitory neurons</b>	<b>25</b>
5.1	Results: linear dependence between mean correlation and motif frequencies .	25
5.2	First theory: second order truncation for network correlation . . . . .	27
5.3	Improved theory: resumming to approximate higher-order contributions to network correlation . . . . .	30
<b>6</b>	<b>Heterogeneous networks</b>	<b>33</b>
6.1	Performance of the homogeneous approximation . . . . .	33
6.2	Heterogeneous theory . . . . .	34
<b>7</b>	<b>Comparisons with IF simulations</b>	<b>36</b>
<b>8</b>	<b>Discussion</b>	<b>37</b>
<b>A</b>	<b>Approximating <math>\rho^{\text{avg}}</math> from <math>\langle \tilde{C}^\infty \rangle</math></b>	<b>42</b>
<b>B</b>	<b>Proof of bound on <math>q_{\text{div}}, q_{\text{con}}, q_{\text{ch}}</math> for one population</b>	<b>43</b>
<b>C</b>	<b>Graph generation methods</b>	<b>43</b>
<b>D</b>	<b>Compensating for fluctuations in empirical connection probability</b>	<b>45</b>
<b>E</b>	<b>Proof of Proposition 4.1</b>	<b>46</b>
<b>F</b>	<b>Proof of Proposition 5.1</b>	<b>49</b>
<b>G</b>	<b>Expression of the linear dependence between <math>\langle \tilde{C}_{EE} \rangle</math> and motifs</b>	<b>50</b>
<b>H</b>	<b>Intuition for why the resumming approach can produce accurate results</b>	<b>51</b>

# 1 Introduction

Neural networks are highly interconnected: a typical neuron in mammalian cortex receives on the order of a thousand inputs. The resulting collective spiking activity is characterized by *correlated* firing of different cells. Such correlations in spiking activity are the focus of a great deal of theoretical and experimental work. This interest arises because correlations can strongly impact the neural coding of information, by introducing redundancy among different (noisy) neurons, allowing noise cancellation effects, or even serving as additional “channels” of information [Zohary et al., 1994, Singer and Gray, 1995, Shadlen and Newsome, 1998b, Panzeri et al., 1999, Abbott and Dayan, 1999, Sompolinsky et al., 2001, Panzeri et al., 2001, Schneidman et al., 2003, Latham and Nirenberg, 2005, Josić et al., 2009, Beck et al., 2011]. Correlations can also help gate the transfer of signals from one brain area to another [Salinas and Sejnowski, 2000, Fries, 2005, Bruno, 2011], and determine the statistical vocabulary of a network in terms of the likelihood that it will produce a particular spike patterns in its repertoire [Schneidman et al., 2006, Shlens et al., 2006].

Thus, the possible effects of correlations on the encoding and transmission of signals are diverse. Making the situation richer still, correlations can depend on characteristics of the “input” signals (or *stimuli*) themselves. The result is an intriguing situation in which there is no simple set of rules that determines the role of correlations – beneficial, detrimental, or neutral – in neural computation. Moreover, despite major progress, the pattern and strength of correlations in neuronal networks *in vivo* remain under debate [Ecker et al., 2010, Cohen and Kohn, 2011]. This demands tools that will let us predict and understand correlations and their impact in specific types of neural circuits.

In this paper, we take one step toward this goal: We determine how simple statistics of network connectivity contribute to the average level of correlations in a neural population. While other factors such as patterns of correlations in upstream neural populations and the dynamical properties of single cells contribute in important ways, we seek to isolate the role of connection statistics in the most direct way possible. We return to the question of how our results on connectivity might be combined with these other factors at several places in the text, and in the Discussion.

We define the statistics of network connectivity via *motifs*, or subgraphs that are the building blocks of complex networks. While there are many ways to characterize connectivity, we choose the motif-based approach for two main reasons. First, we wish to follow earlier work in theoretical neuroscience in which the frequency of small motifs is used to define low-order statistics of network adjacency matrices [Zhao et al., 2011]. Second, recent experiments have characterized the frequencies with which different motifs occur in biological neural networks, and found intriguing deviations from what we would expect in the case of unstructured, random connectivity [Song et al., 2005, Haeusler et al., 2009].

Fig. 1 depicts the motifs we will consider: a single cell projecting to two downstream cells (diverging motif), a single cell receiving inputs from two upstream cells (converging motif), and a series of two connections in and out of a central cell (chain motif). To assess how prevalent these motifs are in a given network, we count their occurrences, and compare the observed motif counts with those expected in a random graph [Song et al., 2005]. This is

a *regular* network which has the same total number of connections, and in which each cell has the same, evenly divided number of incoming and outgoing connections – i.e, the same *in and out degree*. (Importantly, the (relative) motif counts for such regular graphs agree with those in the classical model of random graph, the Erdős-Rényi model in the limit of large network size; thus, when we refer to the prevalence of network motifs, this means in comparison to either a regular or Erdős-Rényi graph.)

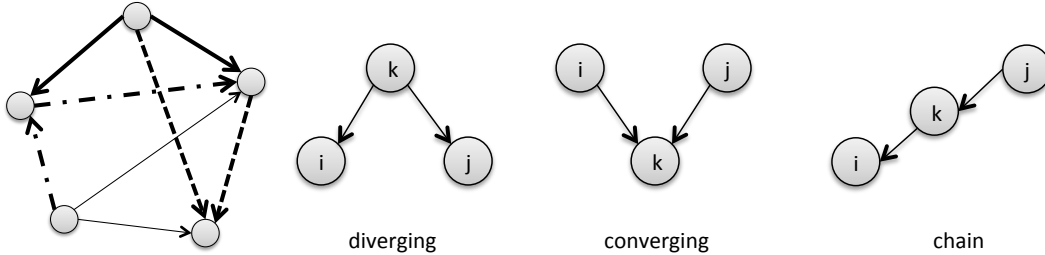


Figure 1: (Left) Counting motifs in a network. Example of a diverging motif in the network is shown in bold solid line. Similarly a converging and a chain motif are shown in dashed line and dash-dotted lines, respectively. In total, the network has 8 connections, 6 diverging, 7 converging and 5 chain motifs. (Right) The different types of second order motifs.

Fig. 2 illustrates the importance of network motifs in determining the average correlation across the network. Here, we simulate 265 different networks of excitatory and inhibitory, exponential integrate and fire cells [Dayan and Abbot, 2001]. Importantly, we bias each neuron so that it fires with the same rate, regardless of its connectivity. We explain in more detail below that this is important to isolate the effects of network connectivity alone. The black dots show the average correlation for Erdős-Rényi networks that have different connection probabilities  $p$ . As expected, correlations increase with connection probability. Next, the grey dots show the average correlation in networks that all have the same connection probability ( $p = 0.2$ ), but with a different prevalence of motifs compared to the corresponding Erdős-Rényi model. Interestingly, the range of correlation values obtained at this fixed connection probability  $p$  is as large as that obtained in the Erdős-Rényi networks over a range of  $p$  values that reaches to roughly twice the connectivity. Thus, motifs – over and above net connectivity – play a strong role in determining the strength of correlations across a network.

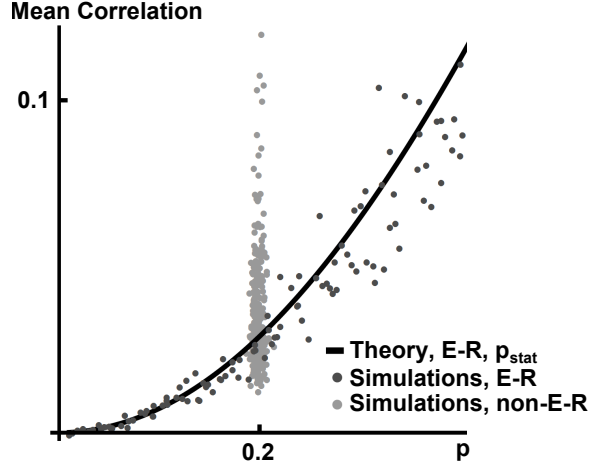


Figure 2: Impact of changing motif frequencies on mean correlation in non-Erdős-Rényi networks (gray dots) compared with the effect of changing connection probability in Erdős-Rényi networks (black dots). Mean correlation coefficients (averaged over all cell pairs in the network) are plotted against connection probability  $p_{\text{stat}}$  in the Erdős-Rényi model. Keeping  $p_{\text{stat}} = 0.2$ , and varying second order motif frequencies strongly affects average correlations. The curve presents the theoretical predictions from Eq. (45). For network parameters see Fig. 3.

But which motifs contribute? And are the second-order motifs of Fig. 1 sufficient, or must higher-order motifs (involving four or more cells) be included to understand average correlations? Fig. 3 suggests part of the answer. Here we used the same 265 network samples represented by the gray dots in Fig. 2. Connection probability was kept at  $p \approx 0.2$ , but frequencies of second order motifs differed. First, panel **A** of Fig. 3 shows the dependence of average network correlations on the total motif counts in the network. Disappointingly, no clear relationship is evident.

A more careful approach is presented in panel **B** of Fig. 3. First, we plot only the mean correlation among cell pairs of a given type – here, excitatory cell pairs. Second, we do not lump together all motifs shown in Fig. 1. Instead, we calculate weighted motif counts that are classified according to the types of constituent neurons. The three panels now exhibit a much clearer trend: Correlation levels vary systematically with weighted motif counts for the diverging and chain motifs, while there is no clear dependence on the converging motif. This improved linear fit is quantified by the coefficient of determination  $R^2$  (as usual, the squared correlation coefficient), for the mean correlation among excitatory cells regressed against the motif counts.

Our goal in the balance of this paper is to explain two aspects of the relationship suggested in panel **B** of Fig. 3. First, we show how to derive analytically a relationship between motif counts and network-wide correlation that successfully identifies these trends. We previously built on the work of Lindner et al. [2005] to derive an explicit expression for the pairwise correlation between cells in a network [Trousdale et al., 2012]; close analogs of this expression have been derived via related approaches for Hawkes processes [Hawkes, 1971b, Pernice et al.,

2011, 2012]. This expression shows how patterns of network connectivity, together with the dynamical properties of single cells, shape correlations. Here we extend our previous work to show how the same expression can reveal the impact of motifs on correlations.

Secondly, we use this analytical approach to obtain a result that is not readily apparent from Fig. 3. We show that – for a range of different network models (see Sec. 3) – the average network correlation can be closely approximated using the connection probability and frequencies of motifs that involve *only two* connections between cells (See Fig. 1). Specifically, we show that the prevalence of diverging and chain motifs tends to increase average correlation (in excitatory only networks), while converging motifs have no effect (in either single population or multi-population networks) .

The paper is structured as follows. We explain in Sec. 2 our setup and introduce the linear response theory which is the basis of our analysis. Sec. 3 defines the three types of motifs and their frequencies, and discusses the classes of networks to which we test and apply our theory. For simplicity, we first demonstrate our analysis in a case of a population of a single population of excitatory cells in Sec. 4, and then generalized to interacting populations of excitatory and inhibitory cells in Sec. 5. We conclude by discussing a crucial assumption of our network model (Sec. 6), and compare our theory with simulations of integrate and fire neuron networks (see Sec. 7). Biological interpretations, limitations, and extensions are covered in our Discussion, and an appendix and table of notation follow.

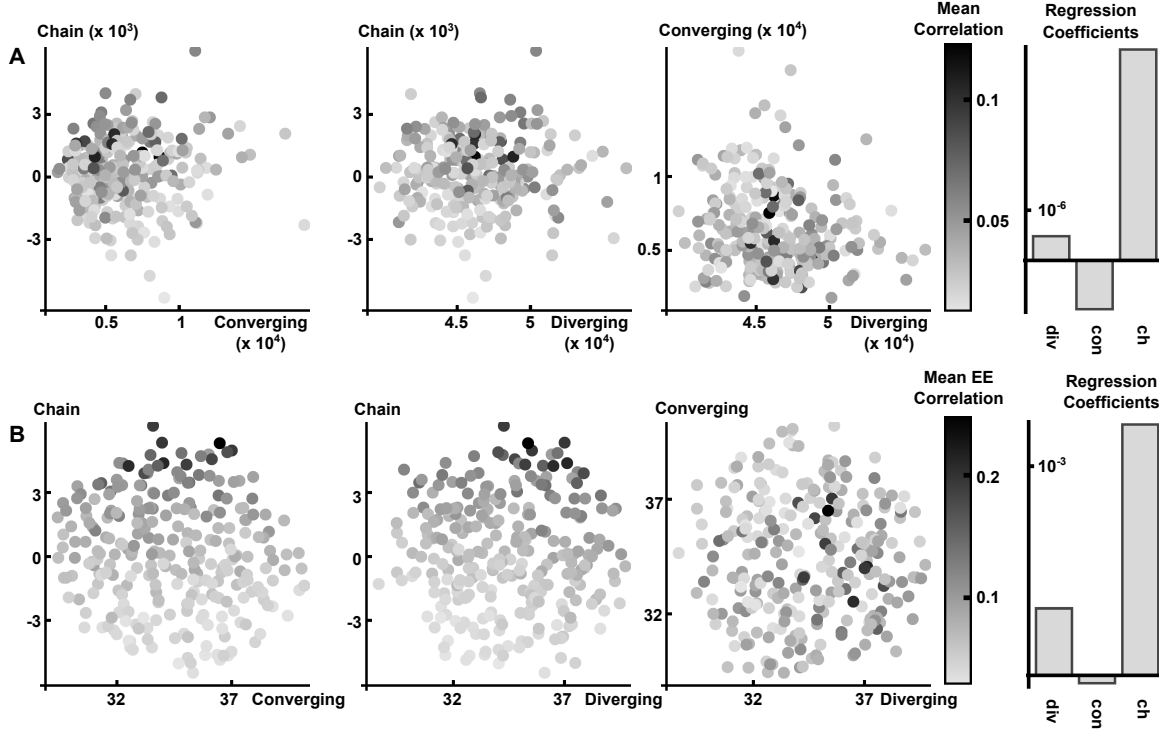


Figure 3: Mean correlation as a function of motif frequency in networks of excitatory and inhibitory neurons. **A** Correlation coefficient averaged over *all* cell pairs as a function of total motif counts. **B** Correlation coefficients averaged over pairs of two excitatory cells, as a function of weighted sums of subgroup motif counts (see Fig. 4 and Eq. (44)). The linear combination coefficients are determined by the resumming theory (see Sec. 5.3). The network consists of 51 excitatory neurons and 49 inhibitory neurons with excitatory and inhibitory coupling strengths set at 22.82 mV·ms or -22.82 mV·ms. All neurons have the same uncoupled cellular dynamic parameters (see Eq. (1)):  $\tau_i = 20, v_{th} = 20, v_r = -60, \tau_{ref} = 2, E_{L,i} = -60, E_i = 8, \sigma_i^2 = 12, v_T = -53, \Delta_T = 1.4, \tau_{S,i} = 5, \tau_{D,i} = 10$ . Spike count correlation coefficients were calculated using a 500ms window size. These graphs are re-sampled from 4096 degree distribution generated graphs using Latin hypercube sampling based on parameter space defined by axis in panel **B** over a reliably sampled region. In panel **A**, some of the motifs are given a “positive” count – i.e., if they involve two excitatory or two inhibitory connections in a chain – and others are given a negative count. For example a converging motif with one excitatory presynaptic cell and one inhibitory presynaptic cell will contribute negatively.

## 2 Neuron models, cross-spectra, and measures of collective network activity

In this section we describe a model spiking network composed of integrate and fire (IF) neurons. We introduce the measures that we will use to quantify the level of correlations

between cell pairs in the network. The analytical approximations of these correlations given in Eq. (6) will be the basis for the subsequent analysis. While we develop the theory for networks of IF neurons, the main ideas are applicable to many other models (e.g. Hawkes model). We return to this point in the Discussion.

## 2.1 Networks of integrate-and-fire neurons

In a network of *integrate-and-fire* (IF) units, the dynamics of each cell are described by a single membrane voltage variable  $v_i$ , which satisfies

$$\tau_i \dot{v}_i = -(v_i - E_{L,i}) + \psi(v_i) + E_i + \sqrt{\sigma_i^2 \tau_i} \xi_i(t) + f_i(t). \quad (1)$$

Here,  $E_{L,i}$  is the leak reversal potential,  $\psi(v_i)$  is a spike generating current, and  $E_i$  is the mean external synaptic input from sources not modeled. In numerical simulations we use the exponential integrate-and-fire (EIF) model [Fourcaud-Trocmé et al., 2003], so that  $\psi(v) \equiv \Delta_T \exp[(v - v_T)/\Delta_T]$ . Each cell has independent fluctuations due to internal noise and external inputs (e.g., from a surrounding network that is not explicitly modeled). We describe such effects by additive terms,  $\sqrt{\sigma_i^2 \tau_i} \xi_i(t)$ , which are Gaussian white noise processes, [White et al., 2000, Renart et al., 2004]. Synaptic input to cell  $i$  from other cells in the network is denoted by  $f_i(t)$  (see below).

When the membrane potential reaches a threshold,  $v_{th}$ , an action potential, or spike, is generated; the membrane potential is then reset to a lower voltage  $v_r$  and held there for an absolute refractory period  $\tau_r$ . We denote the time at which the  $j^{th}$  neuron fires its  $k^{th}$  spike as  $t_{j,k}$ ; taken together, these define this neuron's *spike train*  $y_j(t) = \sum_k \delta(t - t_{j,k})$ . Importantly, synaptic interactions among neurons are initiated at these spike times, so that the total synaptic input to the  $i^{th}$  cell is

$$f_i(t) = \sum_j (\mathbf{J}_{ij} * y_j)(t), \quad \text{where} \quad \mathbf{J}_{ij}(t) = \begin{cases} \mathbf{W}_{ij} \left( \frac{t - \tau_{D,j}}{\tau_{S,j}^2} \right) \exp \left[ -\frac{t - \tau_{D,j}}{\tau_{S,j}} \right] & t \geq \tau_{D,j} \\ 0 & t < \tau_{D,j} \end{cases}.$$

In the absence of a synaptic connection from cell  $j$  to cell  $i$ , we set  $\mathbf{W}_{ij} = 0$ . Hence, the  $N \times N$  matrix of synaptic weights,  $\mathbf{W}$ , defines a directed, weighted network.

In simulations we used the parameters given in the caption of Fig. 3, unless stated otherwise.

## 2.2 Measure of network correlation

Dependencies between the responses of cells in the network may be quantified using the spike train auto- and cross-correlation functions [Gabbiani and Cox, 2010]. For a pair of spike trains  $y_i(t), y_j(t)$ , the cross-covariance function is given by

$$\mathbf{C}_{ij}(\tau) = \mathbf{E} [(y_i(t) - r_i)(y_j(t + \tau) - r_j)],$$



where  $r_i$  and  $r_j$  are the firing rates of the two cells. Here and throughout the paper we assume that the spike trains form a (multivariate) stationary process. The auto-correlation function is the cross-correlation of the output of the cell and itself, and  $\mathbf{C}(t)$  is the matrix of cross-correlation functions. The matrix  $\tilde{\mathbf{C}}(\omega)$  with entries defined by

$$\tilde{\mathbf{C}}_{ij}(\omega) = \mathbf{E} [\tilde{y}_i(\omega) \tilde{y}_j^*(\omega)] ,$$

is the matrix of cross-spectra (see next section). The cross-spectrum of a pair of cells is equivalent to the Fourier transform of their cross-correlation function. [Stratonovich, 1967]

Denote by  $N_{y_i}(t_1, t_2) = \int_{t_1}^{t_2} y_i(s) ds$  the spike count of cell  $i$  over a time window  $[t_1, t_2]$ . The spike count correlation  $\rho_{ij}(T)$  over windows of length  $T$  is defined as

$$\rho_{ij}(T) = \frac{\text{cov} (N_{y_i}(t, t+T), N_{y_j}(t, t+T))}{\sqrt{\text{var} (N_{y_i}(t, t+T)) \text{var} (N_{y_j}(t, t+T))}} .$$

We will make use of the total correlation coefficient  $\rho_{ij}(\infty) = \lim_{T \rightarrow \infty} \rho_{ij}(T)$  which captures dependencies between the processes  $y_i, y_j$  over arbitrarily long timescales, but may also describe well the nature of dependencies over reasonably short timescales [Bair et al., 2001, de la Rocha et al., 2007, Shea-Brown et al., 2008, Rosenbaum and Josić, 2011]. The spike count covariance is related directly to the cross-correlation function by [Tetzlaff et al., 2008]

$$\text{cov} (N_{y_i}(t, t+T), N_{y_j}(t, t+T)) = \int_{-T}^T \mathbf{C}_{ij}(\tau) (T - |\tau|) d\tau .$$

Thus, total correlation may be defined alternatively in terms of the integrated cross-correlations (equivalently, the cross-spectra evaluated at  $\omega = 0$ ):

$$\rho_{ij}(\infty) = \frac{\tilde{\mathbf{C}}_{ij}(0)}{\sqrt{\tilde{\mathbf{C}}_{ii}(0) \tilde{\mathbf{C}}_{jj}(0)}} . \quad (2)$$

Throughout this paper, we will use  $\rho_{ij}(\infty)$  as the measure of correlation and use the above equation to calculate it from the cross-spectra matrix [Pernice et al., 2011]. However, our analysis can be similarly applied to study  $\tilde{\mathbf{C}}_{ij}(\omega)$ , and hence the entire correlation function in time [Trousdale et al., 2012].

### 2.3 Linear response approximation of cell response covariance

Linear response theory [Gabbiani and Cox, 2010, Risken, 1996] can be used to approximate the response of single cells, and the joint response of cells in a network [Lindner and Schimansky-Geier, 2001, Brunel et al., 2001, Lindner et al., 2005, Trousdale et al., 2012]. Consider an IF neuron obeying Eq. (1), but with the mean of the inputs  $f(t)$  absorbed into the constant  $E$ . We denote the remaining, *zero-mean* input by  $\epsilon X(t)$ , so that

$$\tau \dot{v} = -(v - E_L) + \psi(v) + E + \sqrt{\sigma^2 \tau} \xi(t) + \epsilon X(t) . \quad (3)$$

For fixed input fluctuations  $\epsilon X(t)$ , the output spike train will be different for each realization of the noise  $\xi(t)$ , and each initial condition  $v(0)$ . The time-dependent firing rate is obtained by averaging the resulting spike train over noise realizations and a stationary distribution of initial conditions. For all values of  $\epsilon$ , this stationary distribution is taken to be the one obtained when  $\epsilon = 0$ . We denote the resulting averaged firing rate as  $r(t) = \langle y(t) \rangle$ . Linear response theory approximates this firing rate as

$$r(t) = r_0 + (A * \epsilon X)(t),$$

where  $r_0$  is the firing rate in the absence of input ( $\epsilon = 0$ ), and the linear response kernel,  $A(t)$ , characterizes the response to first order in  $\epsilon$ . This approximation is remarkably accurate over a wide range of parameters; for example, see [Ostojic et al., 2009, Richardson, 2009].

Next, we turn to the problem of approximating the output of a cell on a single trial, rather than the average across trials. We denote the Fourier transform of a function  $f$  by  $\tilde{f} = \mathcal{F}(f)$ . However, for spike trains we adopt a convention that  $\tilde{y}_i(\omega)$  is the Fourier transform of the mean subtracted spike train  $y_i(t) - r_i$ . Following [Lindner and Schimansky-Geier, 2001, Lindner et al., 2005, Trousdale et al., 2012] we approximate the spiking output of a cell self-consistently by

$$\tilde{y}_i(\omega) \approx \tilde{y}_i^0(\omega) + \tilde{A}_i(\omega) \left( \sum_j \tilde{\mathbf{J}}_{ij}(\omega) \tilde{y}_j(\omega) \right), \quad (4)$$

where  $\tilde{y}_i^0(\omega)$  is a realization of the output of cell  $i$  in the absence of input. Defining the interaction matrix  $\tilde{\mathbf{K}}$  with entries  $\mathbf{K}_{ij}(t) \equiv (A_i * \mathbf{J}_{ij})(t)$ , we can use Eq. (4) to solve for the vector of Fourier transformed spike train approximations

$$\tilde{\mathbf{y}}(\omega) = (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} \tilde{\mathbf{y}}^0(\omega), \quad (5)$$

and matrix of cross-spectra

$$\begin{aligned} \tilde{\mathbf{C}}(\omega) &\approx \tilde{\mathbf{C}}^\infty(\omega) = (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} \langle \tilde{\mathbf{y}}^0(\omega) \tilde{\mathbf{y}}^{0*}(\omega) \rangle (\mathbf{I} - \tilde{\mathbf{K}}^*(\omega))^{-1} \\ &= (\mathbf{I} - \tilde{\mathbf{A}} \mathbf{W} \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{C}}^0 (\mathbf{I} - \tilde{\mathbf{F}}^* \mathbf{W}^T \tilde{\mathbf{A}}^*)^{-1}. \end{aligned} \quad (6)$$

Here  $\tilde{\mathbf{A}}, \tilde{\mathbf{C}}^0, \tilde{\mathbf{F}}$  are diagonal matrices:  $\tilde{\mathbf{A}}_{ii}(\omega) = \tilde{A}_i(\omega)$  is the linear response of cell  $i$ ,  $\tilde{\mathbf{C}}_{ii}^0(\omega) = \langle \tilde{y}_i^0(\omega) \tilde{y}_i^{0*}(\omega) \rangle$  is its “unperturbed” (i.e., without coupling) power spectrum, and  $\tilde{\mathbf{F}}_{ii}(\omega) = \tilde{F}_i(\omega)$  is the Fourier transform of the synaptic coupling kernel from cell  $i$ . As noted later, our results and analysis will hold at all frequencies and thus can be used to study correlations at all timescales. The weighted connectivity matrix  $\mathbf{W}$ , defines the structure of the network.

To simplify the exposition, we initially assume certain symmetries in the network. For instance, we consider *homogeneous networks* in which cells have identical (unperturbed) power spectra, linear response functions, and synaptic kernels. In this case the diagonal matrices in Eq. (6) act like scalars. We slightly abuse notation in this case, and replace

$\tilde{A}_i(\omega)$  by  $\tilde{A}(\omega)$ , and  $\tilde{C}_i^0(\omega)$  by  $\tilde{C}^0(\omega)$ . This allows us to disentangle the effects of network structure from the effects of neuronal responses on network activity. The resulting cross-spectrum matrix (evaluated at  $\omega = 0$ ) is

$$\tilde{\mathbf{C}}^\infty = \tilde{C}^0(\mathbf{I} - \tilde{\mathbf{A}}\mathbf{W})^{-1}(\mathbf{I} - \tilde{\mathbf{A}}\mathbf{W}^T)^{-1} \quad (7)$$

(note that  $\tilde{F}_i(0) = 1$  by definition). We use this simpler expression in what follows, and return to heterogeneous networks in Sec 6. Since we consider only total correlation, we omit the dependence of the spike count correlation coefficient on window size  $T$ . Additionally, all spectral quantities are evaluated at  $\omega = 0$ , so we also suppress the dependence on  $\omega$ . Finally, we define average network correlation by

$$\rho^{\text{avg}} = \frac{1}{N(N-1)} \sum_{i \neq j}^N \rho_{ij} \quad (8)$$

In subsequent sections, we will examine the average covariance across the network

$$\langle \tilde{\mathbf{C}}^\infty \rangle = \frac{1}{N^2} \sum_{ij}^N \tilde{\mathbf{C}}_{ij}^\infty \quad .$$

This average cannot be directly related to that in Eq. (8), where individual summands are normalized, and diagonal terms are excluded. The motif-based theory we develop predicts  $\langle \tilde{\mathbf{C}}^\infty \rangle$ , and gives no information about the specific entries  $\tilde{\mathbf{C}}_{ij}^\infty$ . However,  $\rho^{\text{avg}}$  can be determined approximately from  $\langle \tilde{\mathbf{C}}^\infty \rangle$  alone. We describe these approximations in Appendix A.

## 2.4 Applicability of linear response theory

Our methods depend on two, related sets of conditions for their validity. First, we take our cells to be driven by a white noise background. This background linearizes the response of the cells to sufficiently weak perturbations, improving the accuracy of the approximation (4); its presence is our first condition.

Second, turning to network effects, we assume that the spectral radius  $\Psi(\tilde{\mathbf{K}}) < 1$ , which gives non-singularity of the approximating processes in Eq. (4) and allows us to make the series expansion we describe in Sec 4.2. In practice, we have found that the linear approximation to correlations will cease to provide an accurate approximation before this occurs, likely owing in part to a failure of the perturbative approximation. Furthermore, the approximation seems to be most accurate at weak interaction strengths as characterized by a small radius of the bulk spectrum of  $\tilde{\mathbf{K}}$ .

For Erdős-Rényi networks, Rajan and Abbott [2006] derived an asymptotic (large  $N$ ) characterization of the spectral radius of the synaptic weight matrix. In particular, for an Erdős-Rényi network consisting of only excitatory cells with synaptic weight  $w$ , there will be a single eigenvalue at  $pNw$  with the remaining eigenvalues distributed uniformly in a circle around the origin of radius  $w\sqrt{p(1-p)N}$ . In networks with both excitatory and inhibitory

populations, there is a single outlier at  $pN_E w_E + pN_I w_I$  and all other eigenvalues will be distributed (non-uniformly) within the circle of radius  $\sqrt{p(1-p)(N_E w_E^2 + N_I w_I^2)}$ . We will use these expressions for the bulk spectrum to quantify the strength of interactions given by the asymptotic spectral radius of  $\tilde{\mathbf{K}} = \tilde{\mathbf{A}}\mathbf{W}$  (referred to as the Erdős-Rényi spectral radius).

We note that we used IF simulations to directly confirm the accuracy of the linear response approximation for excitatory-inhibitory networks in Fig. 14. For a complete discussion of the performance of the linear response theory, see [Trousdale et al., 2012].

### 3 Graphical structure of neuronal networks

Our main goal is to determine how the small-scale statistical structure of directed networks influences the collective dynamics they produce – namely, the strength of spike correlations in networks of model neurons. We will quantify network structure using the probability of connections between pairs and among triplets of cells, organized into *network motifs* [Song et al., 2005, Zhao et al., 2011].

#### 3.1 Network motifs and their frequencies

A motif is a subgraph composed of a small number of cells. We classify motifs according to the number of edges they contain. We begin by considering directed networks composed of identical cells. First order motifs contain one connection and hence come in only one type — two cells with a one-way connection. Second order motifs contain two connections, and therefore involve at most three interacting cells. These motifs come in three types: diverging, converging and chain motifs (See Fig. 1) [Song et al., 2005, Zhao et al., 2011]. (Note that in our definition, a cell can appear twice in the triplet of cells that define a second order motif. For example, the chain motif in Fig. 1 is equivalent to a bidirectionally coupled pair of cells when  $i = j$ .)

We will consider mainly the impact of second order motifs, over and above first order effects. The three motifs shown in Fig. 1 arise naturally in our analysis of correlated spiking activity. In particular, we will show that the *frequency* at which each motif occurs in the network can accurately predict levels of correlation across the network.

We next introduce notation that will allow us to make these ideas precise. Let  $\mathbf{W}^0$  be the adjacency matrix, so that  $\mathbf{W}_{i,j}^0 = 1$  indicates the presence of a directed connection from cell  $j$  to cell  $i$ , and  $\mathbf{W}_{i,j}^0 = 0$  indicates its absence. To quantify the frequency of a motif in a given graph, we first count the total number of times the motif occurs, and divide by the total number of possible occurrences in a graph of that size. For first order motifs this definition gives the *empirical* connection probability,

$$p = \left( \sum_{i,j} \mathbf{W}_{i,j}^0 \right) / N^2.$$

The preponderance of second order motifs is measured in two stages. First, we similarly normalize the motif count. Second, we subtract the value expected in a reference graph.

The resulting expressions are:

$$q_{\text{div}} = \sum_{i,j,k} (\mathbf{W}_{i,k}^0 \mathbf{W}_{j,k}^0) / N^3 - p^2 = \left( \sum_{i,j} (\mathbf{W}^0 \mathbf{W}^{0T})_{i,j} \right) / N^3 - p^2, \quad (9)$$

$$q_{\text{con}} = \left( \sum_{i,j} (\mathbf{W}^{0T} \mathbf{W}^0)_{i,j} \right) / N^3 - p^2, \quad (10)$$

$$q_{\text{ch}} = \left( \sum_{i,j} (\mathbf{W}^0 \mathbf{W}^0)_{i,j} \right) / N^3 - p^2, \quad (11)$$

where  $\mathbf{W}^{0T}$  denotes the transpose of  $\mathbf{W}^0$ . Consider the expression defining  $q_{\text{div}}$ : the sum in the first equality simply counts the total number of connections from one cell ( $k$ ) to two others ( $i$  and  $j$ ), and divides by the total number of possible connections of this type ( $N^3$ ). This can be written as matrix multiplication followed by a sum over all entries  $i, j$ , as shown. In each case we subtract the value  $p^2$ , which corresponds to the frequency of the motif in a regular graph, as well as the asymptotic frequency in an Erdős-Rényi graph as the number of cells,  $N$ , diverges to infinity. Indeed, for Erdős-Rényi graphs any edge is present with chance  $p$ , and any second-order motif requires the presence of two edges. Thus,  $q_{\text{div}}$  corresponds to the propensity for a network to display diverging motifs, over and above expectations from an Erdős-Rényi or regular network. The other measures in Eqns. (9-11) have similar interpretations.

The quantities in Eqns. (9-11) can also be interpreted as empirical measures of covariance [Roxin, 2011, Zhao et al., 2011]. If we denote by  $\mathbf{E}_e[\cdot]$  the empirical average over all entries in a given network, then we may write

$$q_{\text{div}} = \mathbf{E}_e[\mathbf{W}_{i,k}^0 \mathbf{W}_{j,k}^0] - \mathbf{E}_e[\mathbf{W}_{i,k}^0] \mathbf{E}_e[\mathbf{W}_{j,k}^0].$$

Equality is attainable for  $q_{\text{div}}$  and  $q_{\text{con}}$  (but not simultaneously). We also note that the quantities defined in Eqns. (9-11) are not independent (but do have three degrees of freedom). If we first sum over indices  $i, j$  in Eq. (9), we can rewrite the expression in terms of in and out degrees,  $\{d_k^{\text{in}}\}, \{d_k^{\text{out}}\}$ . For example,

$$q_{\text{div}} = \sum_k (d_k^{\text{out}})^2 / N^3 - p^2 = \text{var}(d^{\text{out}}) / N^2 \quad (12)$$

is the scaled sample variance of the out degree across the network. Similarly,

$$q_{\text{con}} = \text{var}(d^{\text{in}}) / N^2, \quad q_{\text{ch}} = \text{cov}(d^{\text{out}}, d^{\text{in}}) / N^2, \quad (13)$$

where  $\text{cov}(\cdot, \cdot)$  denotes sample covariance [Zhao et al., 2011]. Therefore,

$$q_{\text{div}} \geq 0, \quad q_{\text{con}} \geq 0, \quad |q_{\text{ch}}| \leq \sqrt{q_{\text{div}} q_{\text{con}}}. \quad (14)$$

We further show in Appendix B that

$$|q_{\text{div}}|, |q_{\text{con}}|, |q_{\text{ch}}| \leq p(1 - p). \quad (15)$$

Eqns. (14-15) identify the attainable ranges of motif frequencies. In generating our networks, we compare the extent of motif frequencies that we produce using a particular scheme against this maximum possible range (see also Sec. 3.2).

In networks composed of excitatory and inhibitory cells, we can represent interactions between cells using a signed connectivity matrix. Edges emanating from inhibitory neurons are represented by negative entries and those from excitatory neurons by positive entries. In this case, motifs are further subdivided according to their constituent cells. For instance, there are 6 distinct diverging motifs, since if we list all  $2^3$  group types of 3 cells, we see that the motif  $E \leftarrow I \rightarrow I$  is the same as  $I \leftarrow I \rightarrow E$ , and  $E \leftarrow E \rightarrow I$  is the same as  $I \leftarrow E \rightarrow E$ . Similarly, there are 6 distinct converging motifs, and 8 distinct chain motifs, for a total of 20 distinct second order motifs (See Fig. 4).

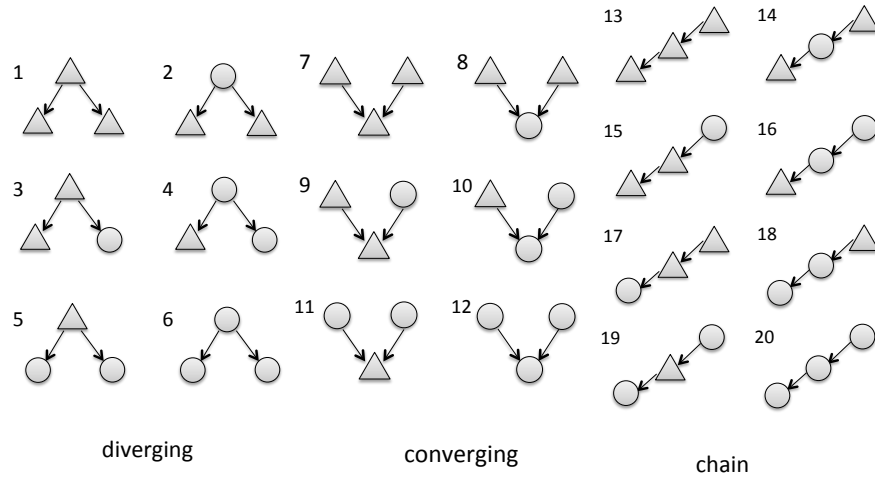


Figure 4: Second order motifs in populations of excitatory and inhibitory cells: There are 20 subtypes of the 3 main motif types. Triangles represent excitatory neurons and circles for inhibitory neurons.

This will clearly lead to some cumbersome notation! Therefore, while populations of interacting E and I cells are our ultimate goal, we first describe our ideas in a population of cells of a single type.

### 3.2 Generating graphs with given motif frequency

To numerically examine the impact of motif frequency on dynamics, we need to generate graphs that are equal in connection probability, but differ in the preponderance of second order motifs. The empirical connection probability in network samples will have small fluctuations around the statistical (i.e. expected) value we fixed. We use two ways of generating such graphs. The first is the degree distribution method [Chung and Lu, 2002] (related to configuration model [Roxin, 2011, Newman, 2003, Newman and Watts, 2001]). Here, following [Zhao et al., 2011] we use a two-sided power law, with various rising and decreasing

exponents, peak locations and truncation limits, as expected in- and out-degree distributions. The other is the *second order network* (SONET) method (for details see [Zhao et al., 2011]). Network samples generated using both methods cover the range of motif frequencies observed experimentally in cortical circuits [Song et al., 2005, Zhao et al., 2011]. Naturally, this experimentally observed range is smaller than the full extent of possibly attainable frequencies (see Eqns. (14-15); however, the SONET method covers this full range as well. Details are given in Appendix C).

We use both methods to generate network samples in the excitatory only case and found similar results; here, we only show data generated using the SONET method as it covers a larger range of motif frequencies. In excitatory-inhibitory networks, we use the degree distribution method.

We emphasize that our approaches below do not *a priori* specify any particular way of generating network samples. However, their accuracy will depend on how this is done. In particular, while we find that our approach is quite accurate for the family of networks that we described above, they can break down in networks that have significant additional structure, a point to which we will return.

## 4 Impact of second order motifs in networks of excitatory cells

As we have shown in Section 2, linear response theory can be used to approximate cross-correlations between cells in a neuronal network. We next explore how the key expression, given in Eq. (7), can be applied to relate the frequency of second order motifs to the average correlation across pairs of cells. For simplicity, we first consider networks consisting only of a single class of excitatory cells. The results extend naturally to the case of two interacting populations of excitatory and inhibitory cells, as we show in Sec. 5.

### 4.1 Results: linear dependence between mean correlation and motif frequencies

Fig. 5 illustrates the relationship between second order motif frequency and average correlation in networks of excitatory cells. In all examples in this section, we use networks of  $N = 100$  cells, with parameters as given in the caption of Fig. 3, and with graph structures generated by the SONET algorithm (See Sec. 3.2). Correlation coefficients between cells are computed using the linear response approximation given in Eq. (6). We find that average correlations depend strongly on the frequency of diverging and especially chain motifs, but only weakly on the frequency of converging motifs. To quantify the linear fit, we use the coefficient of determination between the linear fit and the result of Eq. (6), obtaining  $R^2 = 0.80$ . This high value suggests that the second order motifs are highly predictive of network correlation, in the simplest possible (linear) way (as explained in Appendix D, this prediction can be further improved when we compensate for the fluctuations in empirical connection

probabilities due to the finite size of network samples). In the balance of this section, we explain why this is the case, derive via a resumming theory a nonlinear predictor of mean correlation in terms of motif frequencies, and extract from this an explicit linear relationship between the probability of observing second order motifs and the mean correlation in the network.

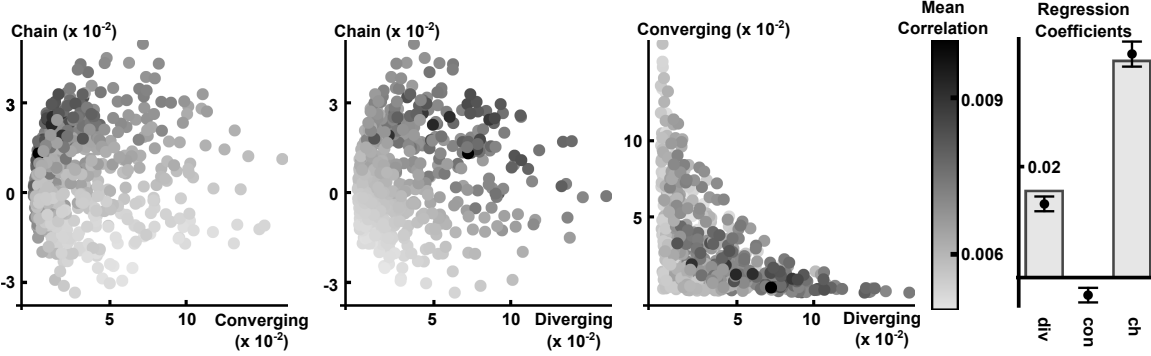


Figure 5: (Left) The relationship between second order motif frequencies and average correlation in purely excitatory networks. Each dot corresponds to a network sample and its shading represents the corresponding average correlation coefficient computed using Eq. (6). Each axis represents one of the three second order motif types as defined in Eq. (9–10). The effective coupling strength, characterized by the Erdős-Rényi spectral radius, was 0.2 for all networks (See Sec. 2.4). (Right) Bars show the linear regression coefficients calculated from the resumming theory (See Eq. (31)) and numerically (points) between motif frequencies and average network correlations. The error bars around each point denote 95% confidence intervals for the regression coefficients. The coefficient of determination  $R^2$  is 0.80. The data was obtained from 512 network samples generated using the SNET algorithm (See Sec. 3.2).

## 4.2 First theory: second order truncation for network correlation

If the spectral radius  $\Psi(\tilde{A}\mathbf{W}) < 1$ , then the matrix inverses in Eq. (7) can be expanded in a power series in  $\tilde{A}\mathbf{W}$  as

$$\frac{\tilde{\mathbf{C}}^\infty}{\tilde{\mathbf{C}}^0} = \sum_{i,j=0}^{\infty} \tilde{A}^{i+j} \mathbf{W}^i (\mathbf{W}^T)^j \quad (16)$$

[Horn and Johnson, 1990]. Terms in this expansion correspond naturally to paths through the graph defining the network [Trousdale et al., 2012, Pernice et al., 2011, Rangan, 2009a,b]. For example, the terms  $(\tilde{A}^2 \mathbf{W}^2)_{ij} = \tilde{A}^2 \sum_k \mathbf{W}_{ik} \mathbf{W}_{kj}$  give the contributions of length two chains between pairs of cells  $i, j$ . Entries in the matrices  $\tilde{A}^3 \mathbf{W}^2 \mathbf{W}^T$ ,  $\tilde{A}^3 \mathbf{W} (\mathbf{W}^T)^2$  give the contributions of diverging motifs of third order consisting of a projection from a single cell to a cell pair. In this case, one branch of the motif is of length two, while the other is a direct connection.



Let  $\mathbf{1}_{N_1 N_2}$  denote the  $N_1 \times N_2$  matrix of ones, and define the  $N$ -vector  $\mathbf{L} = \frac{1}{N} \mathbf{1}_{N,1}$ . We define the orthogonal projection matrices  $\mathbf{H}, \mathbf{\Theta}$  which will play a crucial role in the following analysis:

$$\mathbf{H} = N\mathbf{L}\mathbf{L}^T, \quad \mathbf{\Theta} = \mathbf{I} - \mathbf{H}. \quad (17)$$

Note that if  $\mathbf{X}$  is an  $N \times N$  matrix, then  $\mathbf{L}^T \mathbf{X} \mathbf{L} =: \langle \mathbf{X} \rangle$  is the empirical average of all entries in  $\mathbf{X}$ . We first observe that the empirical network connection probability can be obtained from the adjacency matrix,  $\mathbf{W}^0$ , as

$$p = \mathbf{L}^T \mathbf{W}^0 \mathbf{L}.$$

We can also express second order motif frequencies in terms of intra-network averages. For instance,

$$\begin{aligned} q_{\text{div}} &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^0 \mathbf{W}^{0T} \mathbf{L} - p^2 \\ &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^0 (\mathbf{H} + \mathbf{\Theta}) \mathbf{W}^{0T} \mathbf{L} - p^2 \\ &= (\mathbf{L}^T \mathbf{W}^0 \mathbf{L}) (\mathbf{L}^T \mathbf{W}^{0T} \mathbf{L}) + \frac{1}{N} \mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^{0T} \mathbf{L} - p^2 \\ &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^{0T} \mathbf{L}. \end{aligned} \quad (18)$$

Similarly,  $q_{\text{con}}, q_{\text{ch}}$  may be expressed as

$$\begin{aligned} q_{\text{con}} &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^{0T} \mathbf{W}^0 \mathbf{L} - p^2 \\ &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^{0T} \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}, \end{aligned} \quad (19)$$

and

$$\begin{aligned} q_{\text{ch}} &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^0 \mathbf{W}^0 \mathbf{L} - p^2 \\ &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^{0T} \mathbf{W}^{0T} \mathbf{L} - p^2 \\ &= \frac{1}{N} \mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}. \end{aligned} \quad (20)$$

To relate second-order motif frequencies to mean correlations between pairs of cells, we can truncate Eq. (16) at second order in  $(\tilde{A}\mathbf{W})$ , giving

$$\frac{\tilde{\mathbf{C}}^\infty}{\tilde{\mathbf{C}}^0} \approx I + \tilde{A}w \mathbf{W}^0 + \tilde{A}w \mathbf{W}^{0T} + \left(\tilde{A}w\right)^2 \mathbf{W}^0 \mathbf{W}^{0T} + \left(\tilde{A}w\right)^2 (\mathbf{W}^0)^2 + \left(\tilde{A}w\right)^2 (\mathbf{W}^{0T})^2. \quad (21)$$

To obtain the empirical average of pairwise covariances in the network,  $\langle \tilde{\mathbf{C}}^\infty \rangle$ , we multiply both sides of Eq. (21) on the left and right by  $\mathbf{L}^T$  and  $\mathbf{L}$ , respectively. Making use of Eqns. (18-20), we obtain

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{\mathbf{C}}^0} \approx \frac{1}{N} + 2\tilde{A}wp + 3N \left(\tilde{A}w\right)^2 p^2 + N \left(\tilde{A}w\right)^2 q_{\text{div}} + 2N \left(\tilde{A}w\right)^2 q_{\text{ch}}. \quad (22)$$

How well does this second-order truncation predict levels of correlation across different neural networks? Fig. 6 shows that the truncation correctly captures general trends in levels of correlation from network to network, but makes a substantial systematic error. Here, we plot correlations predicted with the truncated Eq. (22) as an approximation of the full expression (i.e., to all orders) for average correlations given by Eq. (6). Indeed, the truncated expression gives consistent predictions only at very small coupling strengths. We conclude that the terms which were discarded (all terms of order three and higher in  $\tilde{A}\mathbf{W}$ ) can have an appreciable impact on average network correlation, and will next develop methods to capture this impact.

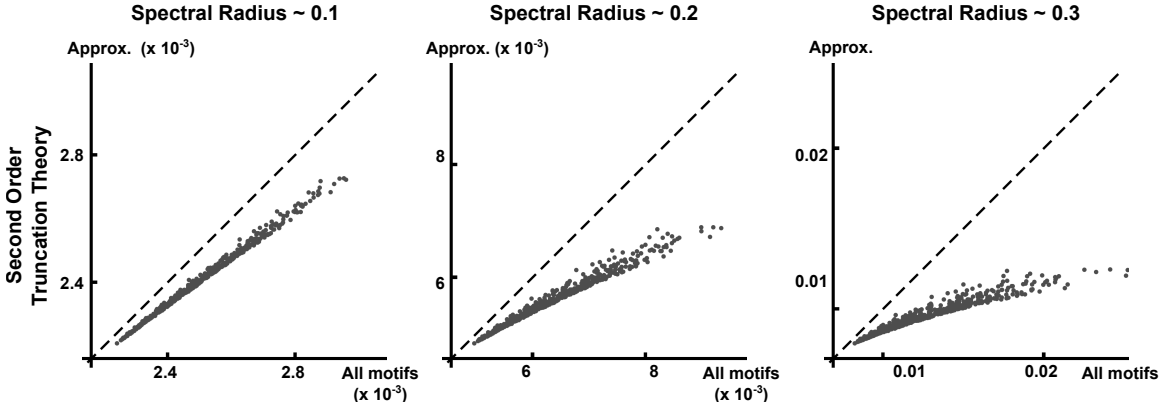


Figure 6: Scatter plot comparing the prediction of average network correlation obtained using Eq. (7) (horizontal axes) to the second order truncation in Eq. (22) (vertical axes). The diagonal line  $y = x$  is plotted for reference. Each panel corresponds to a different coupling strength in the same set of 512 adjacency matrices. The effective coupling strength is characterized by the Erdős-Rényi spectral radius shown at the top of each panel (See Sec. 2.4).

### 4.3 Improved theory: resumming to approximate higher-order contributions to network correlation

A much better approximation of the average network correlation can be obtained by considering the impact of second order motifs on higher order terms in the expansion given by Eq. (16). Note that in an Erdős-Rényi network, every motif of order  $m$  occurs with probability  $p^m$  (with the exception of motifs which involve the same connection multiple times), so that on average  $q_{\text{div}} = q_{\text{con}} = q_{\text{ch}} = 0$ . For non-Erdős-Rényi networks, the expected values of  $q_{\text{div}}$ ,  $q_{\text{con}}$  and  $q_{\text{ch}}$  are typically not zero. As we show next, the introduction of additional second order structure in the network also affects the frequency of motifs of higher order.

Consider the average covariance as determined from the full order linear response approximation of Eq. (16):

$$\frac{\langle \tilde{C}^\infty \rangle}{\tilde{C}^0} = \sum_{i,j=0}^{\infty} \left( \tilde{A}w \right)^{i+j} \mathbf{L}^T (\mathbf{W}^0)^i (\mathbf{W}^{0T})^j \mathbf{L}. \quad (23)$$

We will first express every term in the sum given by Eq. (23) approximately in terms of first and second order motif frequencies ( $p, q_{\text{div}}, q_{\text{con}}$  and  $q_{\text{ch}}$ ). We illustrate this approximation in two examples, before proceeding to the general calculation.

Consider the term  $\mathbf{L}^T (\mathbf{W}^0)^3 \mathbf{L}$  corresponding to the average number of length three chains connecting a pair of cells. Using  $\mathbf{I} = \mathbf{H} + \mathbf{\Theta}$ , we can proceed as in the computation leading to Eq. (18),

$$\begin{aligned} \mathbf{L}^T (\mathbf{W}^0)^3 \mathbf{L} &= \mathbf{L}^T \mathbf{W}^0 (\mathbf{H} + \mathbf{\Theta}) \mathbf{W}^0 (\mathbf{H} + \mathbf{\Theta}) \mathbf{W}^0 \mathbf{L} \\ &= \mathbf{L}^T \mathbf{W}^0 \mathbf{H} \mathbf{W}^0 \mathbf{H} \mathbf{W}^0 \mathbf{L} + \mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{H} \mathbf{W}^0 \mathbf{L} \\ &\quad + \mathbf{L}^T \mathbf{W}^0 \mathbf{H} \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L} + \mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}. \end{aligned} \quad (24)$$

We next replace  $\mathbf{H}$  by  $N\mathbf{L}\mathbf{L}^T$  in the last expression to obtain

$$\begin{aligned} \mathbf{L}^T (\mathbf{W}^0)^3 \mathbf{L} &= N^2 (\mathbf{L}^T \mathbf{W}^0 \mathbf{L}) (\mathbf{L}^T \mathbf{W}^0 \mathbf{L}) (\mathbf{L}^T \mathbf{W}^0 \mathbf{L}) + N (\mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}) (\mathbf{L}^T \mathbf{W}^0 \mathbf{L}) \\ &\quad + N (\mathbf{L}^T \mathbf{W}^0 \mathbf{L}) (\mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}) + \mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}. \end{aligned} \quad (25)$$

Here, the first three terms are composed of factors that correspond to the connection probability ( $\mathbf{L}^T \mathbf{W}^0 \mathbf{L}$ ) and second order chain motif frequency ( $\mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}$ ). These terms provide an estimate of the frequency of a length three chain in a graph, *in terms of the frequency of smaller motifs that form the chain*. The last term,  $\mathbf{L}^T \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{\Theta} \mathbf{W}^0 \mathbf{L}$ , gives the frequency of occurrence of the length three chain in addition to that obtained by chance from second order motifs. Such higher order terms will be gathered separately and denoted by *h.o.t.* in the approximation. We therefore obtain,

$$\mathbf{L}^T (\mathbf{W}^0)^3 \mathbf{L} = N^2 (p^3 + 2pq_{\text{ch}}) + \text{h.o.t.} \quad (26)$$

The exact form of this expression can be understood by referring to Fig. 7: The leading  $N^2$  denotes the number of possible length three chains between a pair of cells (such a chain can pass through  $N^2$  different intermediate pairs of cells) and  $p^3$  represents the probability that one of these length three chains is “present” in an Erdős-Rényi graph. Recall that  $q_{\text{ch}}$  measures the probability that a length two chain is present, above that expected in an Erdős-Rényi network. Therefore,  $pq_{\text{ch}}$  represents a second order estimate of the probability above (Erdős-Rényi) chance that the first two connections in the chain ( $q_{\text{ch}}$ ) and the last ( $p$ ) are present simultaneously. The prefactor 2 appears because this may also occur if the first and final two connections are present simultaneously.

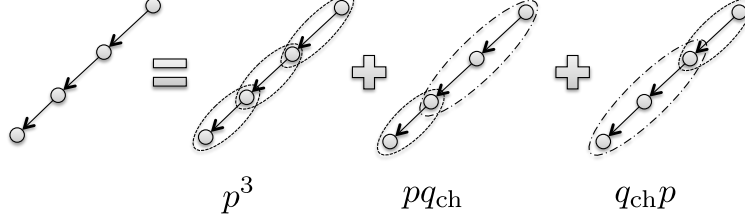


Figure 7: Estimating the number of occurrences of a third order chain motif. The first 3 terms in Eq. (25) correspond to different ways that a chain of length three can be composed from smaller motifs. As explained in the text, each of the three terms represents the estimated probability that smaller motifs occur in one of the arrangements on the right.

As a second example, consider the term  $\mathbf{L}^T (\mathbf{W}^0)^2 (\mathbf{W}^{0T})^2 \mathbf{L}$ , corresponding to indirect diverging motifs with two connections on each branch. A computation similar to that used to obtain Eq. (26) now gives

$$\mathbf{L}^T (\mathbf{W}^0)^2 (\mathbf{W}^{0T})^2 \mathbf{L} = N^3 [p^4 + p^2 q_{\text{div}} + 2p^2 q_{\text{ch}} + q_{\text{ch}}^2] + \text{h.o.t.} \quad (27)$$

The four terms in this sum can be understood with the help of Fig. 8:  $p^4$  represents the chance of observing the motif in an Erdős-Rényi network. The product  $p^2 q_{\text{div}}$  is the estimated probability above (Erdős-Rényi) chance that the motif is formed by two connections emanating from the source cell, present simultaneously and independently with two connections emanating from the tips of the branches. The term  $p^2 q_{\text{ch}}$  gives the estimated probability, above (Erdős-Rényi) chance, that one branch is present, simultaneously and independently, from two connections which form the other branch (the prefactor 2 concerns the probability of this occurring in each of the two branches). The last term,  $q_{\text{ch}}^2$ , gives the estimated probability, above the Erdős-Rényi chance level, that two length two chains simultaneously emanate from the root cell.

In Eq. (27), *h.o.t.* denotes the three distinct terms

$$\begin{aligned} & \mathbf{L}^T \mathbf{W}^0 \Theta \mathbf{W}^0 \Theta \mathbf{W}^{0T} \Theta \mathbf{W}^{0T} \mathbf{L}, \quad N(\mathbf{L}^T \mathbf{W}^0 \mathbf{L})(\mathbf{L}^T \mathbf{W}^0 \Theta \mathbf{W}^{0T} \Theta \mathbf{W}^{0T} \mathbf{L}), \\ & N(\mathbf{L}^T \mathbf{W}^0 \Theta \mathbf{W}^0 \Theta \mathbf{W}^{0T} \mathbf{L})(\mathbf{L}^T \mathbf{W}^{0T} \mathbf{L}). \end{aligned}$$

These terms contain two or more occurrences of  $\Theta$  in one factor, and hence correspond to motifs of higher than second order. In general, a factor which contains  $m$  occurrences of  $\Theta$  will depend on the frequency of a motif of order  $m + 1$ , beyond that which is imposed by the frequency of motifs of order  $m$ .

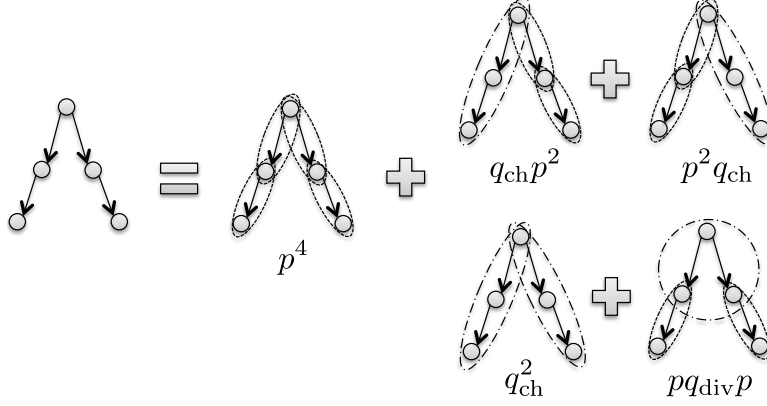


Figure 8: Estimating the number of occurrence of a fourth order motif. Eq. (27) can be understood by decomposing this motif into the constituent first and second order motifs.

The idea behind these two examples extends to all terms in the series in Eq. (23), assuming absolute convergence. Each term in the resulting series contains a factor of the form  $\mathbf{L}^T (\mathbf{W}^0)^i (\mathbf{W}^{0T})^j \mathbf{L}$  corresponding to a motif of order  $i + j$ . This motif corresponds to two chains of length  $i$  and  $j$ , respectively, emanating from the same root cell. To understand the impact of second order motifs we need to decompose this motif as illustrated in Figs. 7 and 8. While this is a challenging combinatorial problem, we show that the answer can be obtained by rearranging the terms in Eq. (23).

Each factor of the form  $\mathbf{L}^T (\mathbf{W}^0)^i (\mathbf{W}^{0T})^j \mathbf{L}$  in Eq. (23) can be split by inserting  $\mathbf{I} = \mathbf{H} + \mathbf{\Theta}$  between each occurrence of  $\mathbf{W}^0$  or  $\mathbf{W}^{0T}$ , as in Eq. (24). The resulting expression can be used to identify the impact of motifs of order  $k$  on terms in the expansion of order  $i + j \geq k$ . The following Proposition, proved in Appendix E, formalizes these ideas.

**Proposition 4.1.** *Let  $\mathbf{H}$  be the rank-1 orthogonal projection matrix generated by the unit  $N$ -vector  $\mathbf{u}$ ,  $\mathbf{H} = \mathbf{u}\mathbf{u}^T$ ,  $\mathbf{\Theta} = \mathbf{I} - \mathbf{H}$ . For any  $N \times N$  matrix  $\mathbf{K}$ , let*

$$\mathbf{K}_n = (\mathbf{K}\mathbf{\Theta})^{n-1} \mathbf{K} = \underbrace{\mathbf{K}\mathbf{\Theta}\mathbf{K} \cdots \mathbf{\Theta}\mathbf{K}}_{n \text{ factors of } \mathbf{K}}.$$

*If the spectral radii  $\Psi(\mathbf{K}) < 1$  and  $\Psi(\mathbf{K}\mathbf{\Theta}) < 1$ , then*

$$\begin{aligned} & \mathbf{u}^T (\mathbf{I} - \mathbf{K})^{-1} (\mathbf{I} - \mathbf{K}^T)^{-1} \mathbf{u} \\ &= \left( 1 - \sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u} \right)^{-1} \left( 1 + \sum_{n,m=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{\Theta} \mathbf{K}_m^T \mathbf{u} \right) \left( 1 - \sum_{m=1}^{\infty} \mathbf{u}^T \mathbf{K}_m^T \mathbf{u} \right)^{-1}. \end{aligned} \quad (28)$$

We will use Prop. 4.1 to derive a relation between second order motif strengths and mean covariances. Assuming that  $\Psi(\tilde{A}w\mathbf{W}^0)$ ,  $\Psi(\tilde{A}w\mathbf{W}^0\mathbf{\Theta}) < 1$ , and setting  $\mathbf{u} = \sqrt{N}\mathbf{L}$  and

$\mathbf{K} = \tilde{A}w\mathbf{W}^0$ , applying Prop. 4.1 to Eq. (7) gives

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0} = \frac{1}{N} \left( 1 - \sum_{n=1}^{\infty} (N\tilde{A}w)^n \mathbf{L}^T \mathbf{W}_n^0 \mathbf{L} \right)^{-1} \left( 1 + \sum_{n,m=1}^{\infty} (N\tilde{A}w)^{n+m} \mathbf{L}^T \mathbf{W}_{n,m}^0 \mathbf{L} \right) \cdot \left( 1 - \sum_{m=1}^{\infty} (N\tilde{A}w)^m \mathbf{L}^T \mathbf{W}_m^{0T} \mathbf{L} \right)^{-1}, \quad (29)$$

where

$$\mathbf{W}_n^0 = \frac{1}{N^{n-1}} \underbrace{\mathbf{W}^0 \Theta \mathbf{W}^0 \dots \Theta \mathbf{W}^0}_{n \text{ factors of } \mathbf{W}^0},$$

$$\mathbf{W}_{n,m}^0 = \frac{1}{N^{n+m-1}} \underbrace{\mathbf{W}^0 \Theta \mathbf{W}^0 \dots \Theta \mathbf{W}^0}_{n \text{ factors of } \mathbf{W}^0} \Theta \underbrace{\mathbf{W}^{0T} \Theta \mathbf{W}^{0T} \dots \Theta \mathbf{W}^{0T}}_{m \text{ factors of } \mathbf{W}^{0T}}.$$

Keeping only terms in Eq. (23) which can be expressed as polynomials of second order in motif frequency and connection probability is equivalent to keeping only terms involving  $\mathbf{W}_1^0$  (connection probability),  $\mathbf{W}_2^0$  (chain motifs) and  $\mathbf{W}_{1,1}^0$  (diverging motifs) in Eq. (29). This yields an expression which involves only first and second order motif frequencies:

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0} = \frac{1}{N} \frac{1 + (N\tilde{A}w)^2 q_{\text{div}}}{\left[ 1 - (N\tilde{A}w) p - (N\tilde{A}w)^2 q_{\text{ch}} \right]^2}. \quad (30)$$

Fig. 9 shows that the approximation to the covariance given by Eq. (30) provides a significant improvement over the second order truncation approximation in Eq. (22) (See Fig. 6). We emphasize that it requires only three scalars that summarize the statistics of the entire connection graph: the overall connection probability and the propensity of two second-order motifs. We offer a heuristic explanation for the effectiveness of the resummation theory based on spectral analysis of  $\mathbf{W}$  in Appendix H.

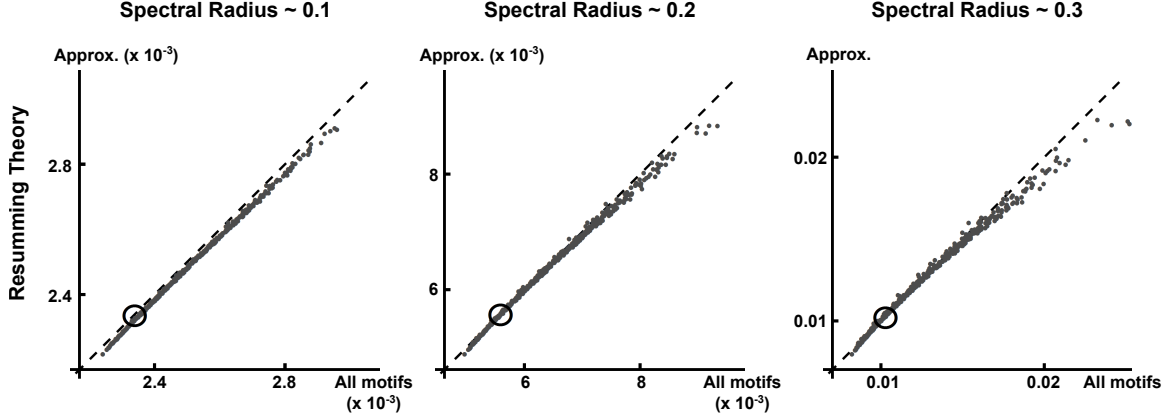


Figure 9: A comparison of the mean correlation obtained using Eq. (7) (horizontal axes) to the resumming approximation in Eq. (30) (vertical axes). The diagonal line  $y = x$  is plotted for reference. Each panel corresponds to a different scaling of coupling strength for the same set of 512 adjacency matrices same as Fig. 6, which are sampled from SOMET (see Sec. 3.2). The effective coupling strength is characterized by the Erdős-Rényi spectral radius, and is recorded at the top of each panel (see Sec. 2.4). Here, the open circle indicates the level of correlation expected from an Erdős-Rényi network with the same overall connection probability and strength.

If we expand the denominator in Eq. (30) as a power series in  $\left[ \left( N \tilde{A} w \right) p + \left( N \tilde{A} w \right)^2 q_{\text{ch}} \right]$ , and keep only terms which are linear in  $q_{\text{div}}$ ,  $q_{\text{ch}}$ , we obtain

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0} = \frac{1}{N(1 - N \tilde{A} w p)^2} + \frac{N(\tilde{A} w)^2}{(1 - N \tilde{A} w p)^2} q_{\text{div}} + \frac{2N(\tilde{A} w)^2}{(1 - N \tilde{A} w p)^3} q_{\text{ch}} + \text{h.o.t.} \quad (31)$$

This linear relation was used to estimate the regression coefficients in Fig. 5 (bar plot).

Finally, we note that similar means may be used to derive versions of Eq. (30) that (unlike Eq. (31)) retain a nonlinear dependence on motif frequencies, but keep motifs of either higher or lower order. To approximate the impact of motifs up to order  $r$ , we would keep terms with factors  $\mathbf{W}_n^0, \mathbf{W}_{n,m}^0$  where  $n, n+m \leq r$  in Eq. (29). For example, if we take  $r = 1$ , we estimate the mean covariance based only on the probability of occurrence of first order motifs – that is, connection probability. This is equivalent to estimating the covariance in idealized Erdős-Rényi networks where  $q_{\text{div}}$  and  $q_{\text{ch}}$ , and their higher order analogs are precisely zero. In Eq. (29), we set all terms involving  $\mathbf{W}$  to zero save  $\mathbf{W}_1^0$ , giving

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0} = \frac{1}{N \left( 1 - N \tilde{A} w p \right)^2}. \quad (32)$$

This predicted mean correlation is indicated by the open dots in Fig. 9 (which accurately describes Erdős-Rényi networks as shown in Fig. 2). This again demonstrates how motif structures exert a large influence over averaged spike correlations.

## 4.4 Correlations in external input

As is natural for communication among different brain areas or layers [Shadlen and Newsome, 1998a], or as arises for certain sensory inputs [Doiron et al., 2004], we next consider the case in which the network under study receives an additional noisy input that is *correlated* from cell-to-cell; in other words, the cells receive common input from upstream sources [de la Rocha et al., 2007, Josić et al., 2009, Trong and Rieke, 2008]. We take this input to have total covariance structure ( $\omega = 0$ )  $\tilde{\mathbf{C}}^\eta = \sigma_X^2 \mathbf{I} + \sigma_X^2 \rho^{\text{input}} (\mathbf{1}_{NN} - \mathbf{I})$ , so that the variance of such external input (not to be confused with the implicitly modeled external noise  $\xi(t)$  term in Eq. (3)) to each neuron is fixed ( $\sigma_X^2$ ) independent of  $\rho^{\text{input}}$  and the correlation coefficient of the inputs to all cell pairs is  $\rho^{\text{input}}$  [Lindner et al., 2005, Marinazzo et al., 2007]. Eq. (7) then has the form (see [Trousdale et al., 2012] for details, and an improvement applicable when the extra noise source is white):

$$\begin{aligned} \tilde{\mathbf{C}}^\infty &= (\mathbf{I} - \tilde{A}\mathbf{W})^{-1} (\tilde{C}^0 \mathbf{I} + \tilde{A}^2 \tilde{\mathbf{C}}^\eta) (\mathbf{I} - \tilde{A}\mathbf{W}^T)^{-1} \\ &= (\mathbf{I} - \tilde{A}\mathbf{W})^{-1} \left[ (\tilde{C}^0 + \tilde{A}^2 \sigma_X^2) \mathbf{I} + (\tilde{A}^2 \sigma_X^2 \rho^{\text{input}}) (\mathbf{1}_{NN} - \mathbf{I}) \right] (\mathbf{I} - \tilde{A}\mathbf{W}^T)^{-1}. \end{aligned}$$

In this case, the output variance in the absence of coupling ( $\mathbf{W} \equiv 0$ ) predicted by the linear response theory is  $(\tilde{C}^0 + \tilde{A}^2 \sigma_X^2)$ . Normalizing and multiplying by  $\mathbf{L}^T, \mathbf{L}$  to arrive at an approximation of average correlation, and applying the ideas of Sec. 4.3, we find that

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0 + \tilde{A}^2 \sigma_X^2} = \frac{B_1 [1 + (N\tilde{A}w)^2 q_{\text{div}}] + NB_2}{N [1 - (N\tilde{A}w)p - (N\tilde{A}w)^2 q_{\text{ch}}]^2},$$

where

$$B_1 = \frac{\tilde{C}^0 + \tilde{A}^2 \sigma_X^2 (1 - \rho^{\text{input}})}{\tilde{C}^0 + \tilde{A}^2 \sigma_X^2}, \quad B_2 = \frac{\tilde{A}^2 \sigma_X^2 \rho^{\text{input}}}{\tilde{C}^0 + \tilde{A}^2 \sigma_X^2}$$

(compare Eq. (30)). We might ask how the presence of input correlations affects output correlations by calculating the *change in output correlation* resulting from input correlations of size  $\rho^{\text{input}}$ :

$$\begin{aligned} \Delta \rho^{\text{output}} &= \left. \frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0 + \tilde{A}^2 \sigma_X^2} - \frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0 + \tilde{A}^2 \sigma_X^2} \right|_{\rho^{\text{input}}=0} \\ &= \frac{B_2 [N - 1 - (N\tilde{A}w)^2 q_{\text{div}}]}{N [1 - (N\tilde{A}w)p - (N\tilde{A}w)^2 q_{\text{ch}}]^2}. \end{aligned} \tag{33}$$

For simplicity, we can linearize this expression in  $q_{\text{div}}, q_{\text{ch}}$  to examine the interaction between second order motifs and input correlations, giving

$$\Delta \rho^{\text{output}} = \frac{B_2}{N} \left[ \frac{N-1}{[1 - (N\tilde{A}w)p]^2} - \frac{(N\tilde{A}w)^2}{[1 - (N\tilde{A}w)p]^2} q_{\text{div}} + \frac{2(N-1)(N\tilde{A}w)^2}{[1 - (N\tilde{A}w)p]^3} q_{\text{ch}} \right]. \tag{34}$$



If we assume that  $w \sim \mathcal{O}(\frac{1}{N})$  (so that the Erdős-Rényi spectral radius  $N\tilde{A}w \sim \mathcal{O}(1)$ , see Sec. 2.4), then, asymptotically,  $q_{\text{div}}$  only has an order  $\mathcal{O}(\frac{1}{N})$  (negative) contribution, while  $q_{\text{ch}}$  has a order  $\mathcal{O}(1)$  contribution to  $\Delta\rho^{\text{output}}$  (note that the first term in (34) is also  $\mathcal{O}(1)$ , which represents the “base” response to correlated input in a Erdős-Rényi network). This implies that in large networks, the chain motif is the most important motif in determining how input correlations will be transferred into network correlations – which will be “output” to the next area downstream.

## 5 Impact of second order motifs in networks of excitatory and inhibitory neurons

Biological neuronal networks are composed of excitatory and inhibitory neurons (EI networks). To treat this case, we next show how our theory extends to the case of networks composed of two interacting subpopulations.

### 5.1 Results: linear dependence between mean correlation and motif frequencies

As in the previous section, we start by a numerical exploration of the contribution of motifs to network-averaged correlation in excitatory-inhibitory networks. We generated 512 networks using the degree distribution method, as described in Sec. 3.2. We evaluated network-averaged correlations using the full linear response expression given in Eq. (7). We then performed a linear regression analysis of the dependence of network correlations on motif frequency in networks of excitatory and inhibitory neurons, for the 20 second order motifs shown in Fig. 4.

The linear regression gives a reasonable fit ( $R^2 \geq 0.69$ , see caption). Chain and diverging motifs contribute significantly to the network-averaged correlation, with some motifs having a higher effect than others. Moreover, mean correlations depend only weakly on converging motifs. Note also that these regression coefficients correspond to how we devise the axes in Fig. 3B. There, we use these regression coefficients (or actually the theory predication for these values based on Eq. (44)) as weights to linearly combine multiple motif frequencies. Specifically, we take linear combinations of the 6 diverging, 6 converging and 8 chain motif frequencies (see Fig. 4) respectively, giving the 3 axes in Fig. 3B.

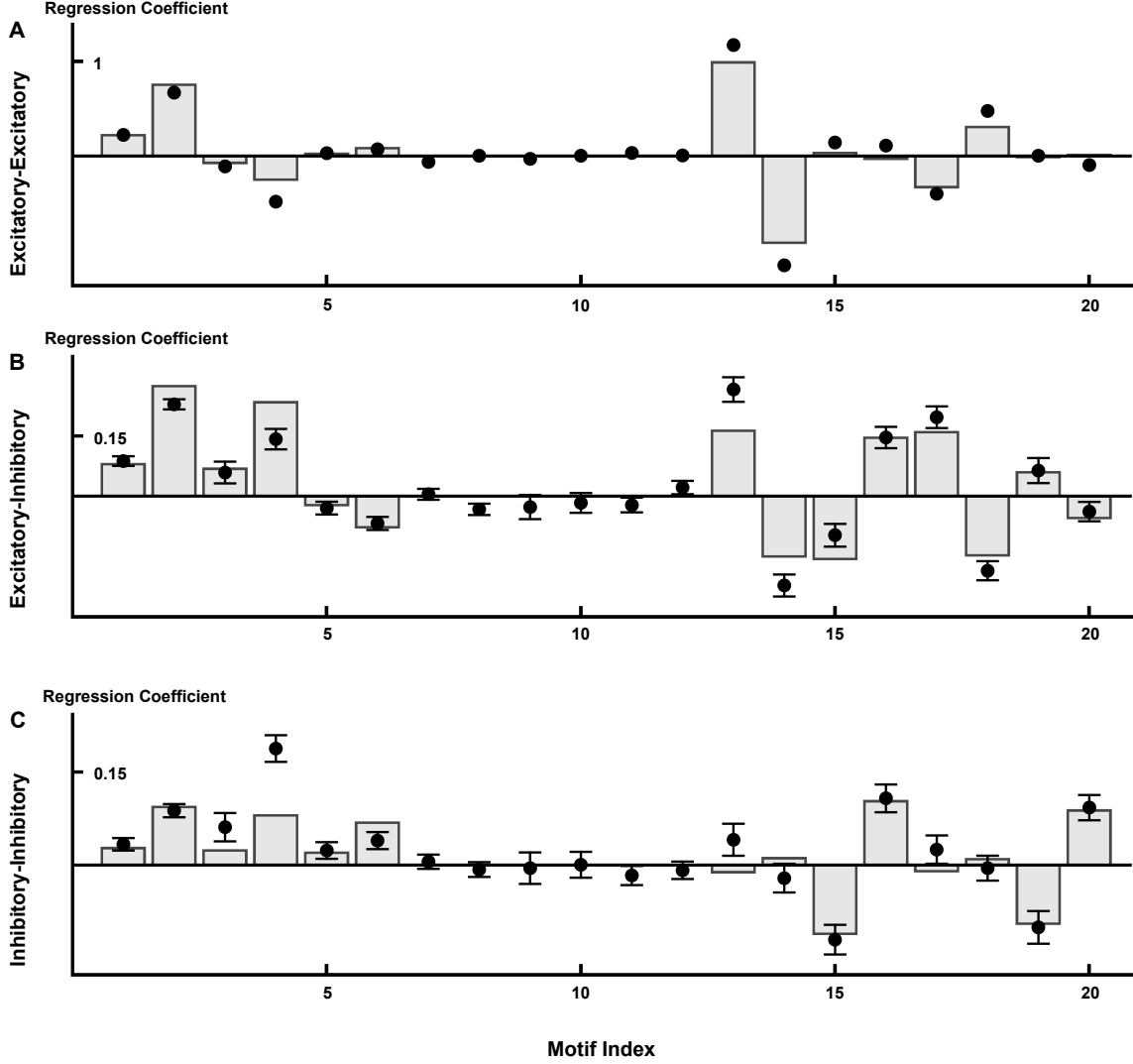


Figure 10: Linear regression coefficients between network-averaged correlations within different subpopulations (vertical axes) and motif frequencies (horizontal axes). The three panels correspond to averages across EE, EI and II cell pairs, from top to bottom. The linear regression coefficient between average correlation and 20 motif frequencies are represented by a dot (See Fig. 4 for the enumerated list of second order motifs in an EI network). Error bars represent a 95% confidence intervals. The vertical bars were obtained from the resummung theory described in Sec. 5.3. Here  $R^2 = 0.91, 0.86, 0.69$  respectively. The spectral radius of  $\tilde{A}\mathbf{W}$  under the Erdős-Rényi assumption is 0.25, and  $N_E = 80$ ,  $N_I = 20$ ,  $|w_I|/|w_E| = 3.707$  so that  $pN_E w_E + pN_I w_I \approx 0$  giving approximate balance between average excitatory and inhibitory inputs. Parameters are given in Fig. 3.

## 5.2 First theory: second order truncation for network correlation

To extend the theory developed in Sec. 4 to EI networks, we need to take into account distinct second order motifs. For instance, there are eight different types of three-cell chains (See Fig. 4). Although this makes the notation more burdensome, the main ideas are the same. Indeed, following a similar approach, the theory can be extended to an arbitrary number of subpopulations.

Consider a network of size  $N = N_E + N_I$ , consisting of  $N_E$  excitatory and  $N_I$  inhibitory neurons. Excitatory (resp. inhibitory) connections have weight  $w_E$  (resp.  $w_I$ ), so that  $w_E > 0$  and  $w_I < 0$ . The connection probability from class  $X$  to class  $Y$  is  $p_{YX}$ . Note that the connectivity matrix can now be written as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{EE} & \mathbf{W}_{EI} \\ \mathbf{W}_{IE} & \mathbf{W}_{II} \end{pmatrix} = \begin{pmatrix} w_E \mathbf{W}_{EE}^0 & w_I \mathbf{W}_{EI}^0 \\ w_E \mathbf{W}_{IE}^0 & w_I \mathbf{W}_{II}^0 \end{pmatrix},$$

where,  $\mathbf{W}_{XY}$  and  $\mathbf{W}_{XY}^0$  are respectively, the weighted and unweighted connection matrices between cells of class  $Y$  to cells of class  $X$ . The population sizes, weights, connection probabilities, and firing rates together determine the balance between excitatory and inhibitory inputs that cells receive (see Discussion). Throughout this section we use networks with same parameters and architecture as in Fig. 10.

First, define the  $N \times 2$  block-averaging matrix  $\mathbf{L}$  by

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_E & 0 \\ 0 & \mathbf{L}_I \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{N_E,1}/N_E & 0 \\ 0 & \mathbf{1}_{N_I,1}/N_I \end{pmatrix},$$

where  $\mathbf{1}_{NM}$  is the  $N \times M$  matrix of ones. We define the two-population analogs of the orthogonal projection matrices  $\mathbf{H}$  and  $\mathbf{\Theta}$  by

$$\mathbf{H} = \mathbf{L} \mathbf{D}_2 \mathbf{L}^T, \quad \text{and} \quad \mathbf{\Theta} = \mathbf{I} - \mathbf{H}, \quad (35)$$

where  $\mathbf{D}_2$  is the  $2 \times 2$  matrix

$$\mathbf{D}_2 = \begin{pmatrix} N_E & 0 \\ 0 & N_I \end{pmatrix}.$$

Direct matrix multiplication shows that if  $\mathbf{X}$  is a matrix with block form

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{EE} & \mathbf{X}_{EI} \\ \mathbf{X}_{IE} & \mathbf{X}_{II} \end{pmatrix},$$

where  $\mathbf{X}_{YZ}$  is an  $N_Y \times N_Z$  matrix, then  $\mathbf{L}^T \mathbf{X} \mathbf{L}$  is a  $2 \times 2$  matrix of block-wise averages of  $\mathbf{X}$ , that is,

$$\langle \mathbf{X} \rangle_B \stackrel{\text{def}}{=} \mathbf{L}^T \mathbf{X} \mathbf{L} = \begin{pmatrix} \mathbf{L}_E^T \mathbf{X}_{EE} \mathbf{L}_E & \mathbf{L}_E^T \mathbf{X}_{EI} \mathbf{L}_I \\ \mathbf{L}_I^T \mathbf{X}_{IE} \mathbf{L}_E & \mathbf{L}_I^T \mathbf{X}_{II} \mathbf{L}_I \end{pmatrix} = \begin{pmatrix} \langle \mathbf{X}_{EE} \rangle & \langle \mathbf{X}_{EI} \rangle \\ \langle \mathbf{X}_{IE} \rangle & \langle \mathbf{X}_{II} \rangle \end{pmatrix}.$$

We will make use of the *empirical average connection strength* matrix  $\mathbf{M}$  given by

$$\mathbf{M} = \mathbf{L}^T \mathbf{W} \mathbf{L} = \begin{pmatrix} w_E p_{EE} & w_I p_{EI} \\ w_E p_{IE} & w_I p_{II} \end{pmatrix}.$$

To examine the dependence of mean correlations *within a block* on the frequency of second order motifs, we consider the block-wise average of the covariance matrix  $\tilde{\mathbf{C}}^\infty$ ,

$$\langle \tilde{\mathbf{C}}^\infty \rangle_B = \mathbf{L}^T \tilde{\mathbf{C}}^\infty \mathbf{L} = \begin{pmatrix} \langle \tilde{\mathbf{C}}_{EE}^\infty \rangle & \langle \tilde{\mathbf{C}}_{EI}^\infty \rangle \\ \langle \tilde{\mathbf{C}}_{IE}^\infty \rangle & \langle \tilde{\mathbf{C}}_{II}^\infty \rangle \end{pmatrix}.$$

Excitatory and inhibitory connection weights need not be equal. It is therefore necessary to consider motif frequencies simultaneously with connection weights. For instance, the contributions of a length two chain passing through an excitatory or inhibitory intermediary cell, such as motifs 13 and 14 in Fig. 4, are not necessarily equal and opposite in sign. Their contributions are dependent on the ratio  $w_E/w_I$ . To account for this, we define motif *strength matrices*  $\mathbf{Q}_{\text{div}}$ ,  $\mathbf{Q}_{\text{con}}$ , and  $\mathbf{Q}_{\text{ch}}$  as follows.

The strength of diverging motifs expected in an Erdős-Rényi network is given by

$$\begin{aligned} \mathbf{Q}_{\text{div}}^{\text{ER}} &= \mathbf{M} \mathbf{D}_2 \mathbf{M}^T = (\mathbf{L}^T \mathbf{W} \mathbf{L}) \mathbf{D}_2 (\mathbf{L}^T \mathbf{W}^T \mathbf{L}) \\ &= \mathbf{L}^T \mathbf{W} \mathbf{H} \mathbf{W}^T \mathbf{L} \\ &= \begin{pmatrix} N_E (w_E p_{EE})^2 + N_I (w_I p_{EI})^2 & N_E w_E^2 p_{EE} p_{IE} + N_I w_I^2 p_{EI} p_{II} \\ N_E w_E^2 p_{EE} p_{IE} + N_I w_I^2 p_{EI} p_{II} & N_E (w_E p_{IE})^2 + N_I (w_I p_{II})^2 \end{pmatrix}. \end{aligned}$$

Here multiplication by  $\mathbf{D}_2$  converts average individual motif strength (e.g.  $w_E^2 p_{EE} p_{IE}$ ) to average total motif strength. The empirical average of the strength of diverging motifs is given by

$$\begin{aligned} \mathbf{Q}_{\text{div}}^{\text{total}} &= \mathbf{L}^T \mathbf{W} \mathbf{W}^T \mathbf{L} \\ &= \begin{pmatrix} \langle \mathbf{W}_{EE} \mathbf{W}_{EE}^T \rangle + \langle \mathbf{W}_{EI} \mathbf{W}_{EI}^T \rangle & \langle \mathbf{W}_{EE} \mathbf{W}_{IE}^T \rangle + \langle \mathbf{W}_{EI} \mathbf{W}_{II}^T \rangle \\ \langle \mathbf{W}_{IE} \mathbf{W}_{EE}^T \rangle + \langle \mathbf{W}_{II} \mathbf{W}_{EI}^T \rangle & \langle \mathbf{W}_{IE} \mathbf{W}_{IE}^T \rangle + \langle \mathbf{W}_{II} \mathbf{W}_{II}^T \rangle \end{pmatrix}. \end{aligned}$$

Hence, we can write the expected total strength of diverging motifs in excess of that expected in an Erdős-Rényi network as

$$\begin{aligned} \mathbf{Q}_{\text{div}} &= \begin{pmatrix} Q_{\text{div}}^{EE} & Q_{\text{div}}^{EI} \\ Q_{\text{div}}^{IE} & Q_{\text{div}}^{II} \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{Q}_{\text{div}}^{\text{total}} - \mathbf{Q}_{\text{div}}^{\text{ER}} \\ &= \begin{pmatrix} N_E w_E^2 q_{\text{div}}^{EE,E} + N_I w_I^2 q_{\text{div}}^{EE,I} & N_E w_E^2 q_{\text{div}}^{EI,E} + N_I w_I^2 q_{\text{div}}^{EI,I} \\ N_E w_E^2 q_{\text{div}}^{EI,E} + N_I w_I^2 q_{\text{div}}^{EI,I} & N_E w_E^2 q_{\text{div}}^{II,E} + N_I w_I^2 q_{\text{div}}^{II,I} \end{pmatrix} \\ &= \mathbf{L}^T \mathbf{W} \mathbf{W}^T \mathbf{L} - \mathbf{L}^T \mathbf{W} \mathbf{H} \mathbf{W}^T \mathbf{L} \\ &= \mathbf{L}^T \mathbf{W} (\mathbf{H} + \boldsymbol{\Theta}) \mathbf{W}^T \mathbf{L} - \mathbf{L}^T \mathbf{W} \mathbf{H} \mathbf{W}^T \mathbf{L} \\ &= \mathbf{L}^T \mathbf{W} \boldsymbol{\Theta} \mathbf{W}^T \mathbf{L}. \end{aligned} \tag{36}$$

where

$$q_{\text{div}}^{XY,Z} = \frac{1}{N_Z} \langle \mathbf{W}_{XZ}^0 \mathbf{W}_{YZ}^{0T} \rangle - p_{XZ} p_{YZ}$$

represents the probability of observing a diverging motif to cells of classes  $X, Y$  from a cell of class  $Z$  in excess of that expected in an Erdős-Rényi network.

It is important to note that the matrix  $\mathbf{Q}_{\text{div}}$  contains motif *strengths* in excess of what would be expected in an Erdős-Rényi network (i.e., number of occurrences, scaled by connection weights and probability of occurrence), while the scalars  $q_{\text{div}}$  still correspond to probabilities.

The strengths and frequencies of converging motifs can be expressed similarly, giving

$$\begin{aligned}\mathbf{Q}_{\text{con}} &= \begin{pmatrix} Q_{\text{con}}^{EE} & Q_{\text{con}}^{EI} \\ Q_{\text{con}}^{IE} & Q_{\text{con}}^{II} \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{L}^T \mathbf{W}^T \mathbf{W} \mathbf{L} - \mathbf{M}^T \mathbf{D}_2 \mathbf{M} = \mathbf{L}^T \mathbf{W}^T \mathbf{\Theta} \mathbf{W} \mathbf{L} \\ &= \begin{pmatrix} N_E w_E^2 q_{\text{con}}^{EE,E} + N_I w_E^2 q_{\text{con}}^{EE,I} & N_E w_E w_I q_{\text{con}}^{EI,E} + N_I w_E w_I q_{\text{con}}^{EI,I} \\ N_E w_E w_I q_{\text{con}}^{EI,E} + N_I w_E w_I q_{\text{con}}^{EI,I} & N_E w_I^2 q_{\text{con}}^{II,E} + N_I w_I^2 q_{\text{con}}^{II,I} \end{pmatrix},\end{aligned}\quad (37)$$

where

$$q_{\text{con}}^{XY,Z} = \frac{1}{N_Z} \langle \mathbf{W}_{XZ}^{0T} \mathbf{W}_{YZ}^0 \rangle - p_{ZX} p_{ZY},$$

represents the probability of observing a converging motif to cells of class  $Z$  from cells of classes  $X, Y$  in excess of that expected in an Erdős-Rényi network.

Finally, for chain motifs,

$$\begin{aligned}\mathbf{Q}_{\text{ch}} &= \begin{pmatrix} Q_{\text{ch}}^{EE} & Q_{\text{ch}}^{EI} \\ Q_{\text{ch}}^{IE} & Q_{\text{ch}}^{II} \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{L}^T \mathbf{W}^2 \mathbf{L} - \mathbf{M} \mathbf{D}_2 \mathbf{M} = \mathbf{L}^T \mathbf{W} \mathbf{\Theta} \mathbf{W} \mathbf{L} \\ &= \begin{pmatrix} N_E w_E^2 q_{\text{ch}}^{EEE} + N_I w_E w_I q_{\text{ch}}^{EIE} & N_E w_E w_I q_{\text{ch}}^{IEE} + N_I w_I^2 q_{\text{ch}}^{IIE} \\ N_E w_E^2 q_{\text{ch}}^{EEI} + N_I w_E w_I q_{\text{ch}}^{EII} & N_E w_E w_I q_{\text{ch}}^{IEI} + N_I w_I^2 q_{\text{ch}}^{III} \end{pmatrix},\end{aligned}\quad (38)$$

where

$$q_{\text{ch}}^{ZYX} = \frac{1}{N_Y} \langle \mathbf{W}_{ZY}^0 \mathbf{W}_{YX}^0 \rangle - p_{ZY} p_{YX}$$

represents the probability of observing a length two chain motif beginning at a cell of type  $X$  and terminating at a cell of type  $Z$ , passing through a cell of type  $Y$ , in excess of what would be expected in an Erdős-Rényi network.

A truncation of Eq. (16) at second order gives an initial approximation of the block-wise average  $\langle \tilde{\mathbf{C}}(0) \rangle_B$ :

$$\begin{aligned}\langle \tilde{\mathbf{C}}^\infty \rangle_B / \tilde{\mathbf{C}}^0 &= \mathbf{L}^T \left[ \mathbf{I} + \tilde{A} (\mathbf{W} + \mathbf{W}^T) + \tilde{A}^2 (\mathbf{W}^2 + \mathbf{W}^{2T} + \mathbf{W} \mathbf{W}^T) \right] \mathbf{L} + \text{h.o.t.} \\ &\approx \left[ \mathbf{L}^T \mathbf{L} + \tilde{A} (\mathbf{M} + \mathbf{M}^T) + \tilde{A}^2 (\mathbf{M} \mathbf{D}_2 \mathbf{M} + \mathbf{M}^T \mathbf{D}_2 \mathbf{M}^T + \mathbf{M} \mathbf{D}_2 \mathbf{M}^T) \right. \\ &\quad \left. + \tilde{A}^2 (\mathbf{Q}_{\text{ch}} + \mathbf{Q}_{\text{ch}}^T + \mathbf{Q}_{\text{div}}) \right].\end{aligned}\quad (39)$$

Terms not involving a second order motif matrix correspond to the contributions of motifs up to second order in a purely Erdős-Rényi network — for example,  $\mathbf{M} \mathbf{D}_2 \mathbf{M}$  gives the expected contribution of length two chains in an Erdős-Rényi network.

Fig. 11 compares the second order truncation given by Eq. (39) with the mean correlations obtained from the entire series in Eq. (16). The correlations in the network of inhibitory and excitatory cells can be appreciable, even when the spectral radius of matrix  $\tilde{A} \mathbf{W}$  is

much smaller than one. However, the second order truncation of Eq. (39) gives a poor approximation of network-averaged correlations.

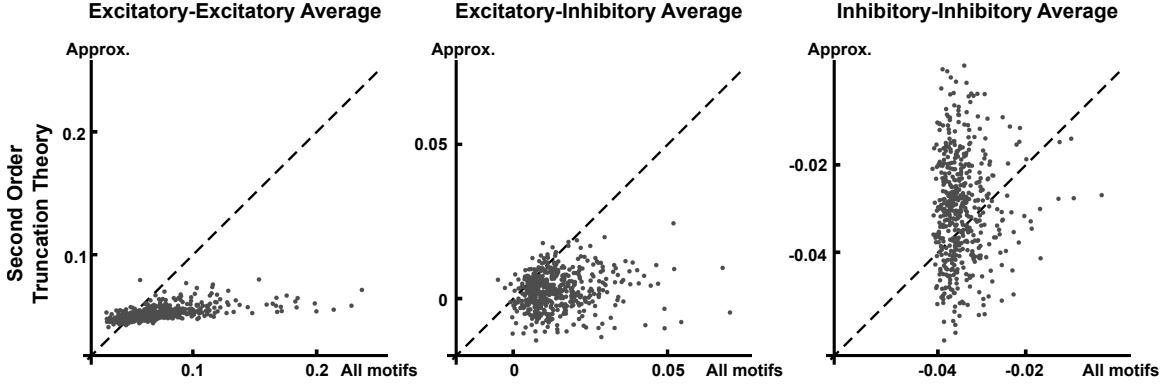


Figure 11: Block-wise average correlations obtained from the second order truncation of Eq. (39). On the horizontal axis are the average correlation between cells from the given classes (calculated from Eq. (7)), and the approximate value obtained from the truncation is given on the vertical axis. The diagonal line  $y = x$  corresponds to perfect agreement between the true value and the approximation. The spectral radius of  $\tilde{A}\mathbf{W}$  under the Erdős-Rényi assumption is 0.33 (see Sec. 2.4). Other network parameters are the same as in Fig. 10, 12 and 13. Here  $R^2 = 0.33, 0.02, 0.0005$  respectively for the three panels, confirming that the truncation approach gives a poor prediction of network correlation. Here, the open circle indicates the level of correlation expected from an Erdős-Rényi network with the same overall connection probabilities and strengths.

### 5.3 Improved theory: resumming to approximate higher-order contributions to network correlation

As in the case of a single population, we can improve our prediction of mean correlations by accounting for the contributions of second order motifs to all orders in connection strength. The equivalent of Eq. (23) has the form

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle_B}{\tilde{C}^0} = \frac{1}{\tilde{C}^0} \mathbf{L}^T \tilde{\mathbf{C}}^\infty \mathbf{L} = \sum_{i,j=0}^{\infty} \tilde{A}^{i+j} \mathbf{L}^T \mathbf{W}^i (\mathbf{W}^T)^j \mathbf{L}, \quad (40)$$

where  $\mathbf{W}$  and  $\mathbf{L}$  are as defined in the previous section.

First, we generalize Prop. 4.1 to the case of two populations (for a full proof, see Appendix F)

**Proposition 5.1.** *Let  $\mathbf{H} = \mathbf{U}\mathbf{U}^T$  be a orthogonal projection matrix generated by a  $N \times M$  matrix  $\mathbf{U}$ , whose columns are orthonormal vectors. Define  $\mathbf{\Theta} = \mathbf{I} - \mathbf{H}$ . For any  $N \times N$  matrix  $\mathbf{K}$ , define  $\mathbf{K}_n$  as*

$$\mathbf{K}_n = \underbrace{\mathbf{K}\mathbf{\Theta}\mathbf{K} \cdots \mathbf{\Theta}\mathbf{K}}_{n \text{ factors of } \mathbf{K}}.$$

If spectral radius  $\Psi(\mathbf{K})$ ,  $\Psi(\mathbf{K}\Theta) < 1$ , we have

$$\begin{aligned} & \mathbf{U}^T(\mathbf{I} - \mathbf{K})^{-1}(\mathbf{I} - \mathbf{K}^T)^{-1}\mathbf{U} \\ &= \left( \mathbf{I} - \sum_{n=1}^{\infty} \mathbf{U}^T \mathbf{K}_n \mathbf{U} \right)^{-1} \left( \mathbf{I} + \sum_{n,m=1}^{\infty} \mathbf{U}^T \mathbf{K}_n \Theta \mathbf{K}_m^T \mathbf{U} \right) \left( \mathbf{I} - \sum_{m=1}^{\infty} \mathbf{U}^T \mathbf{K}_m^T \mathbf{U} \right)^{-1}. \end{aligned} \quad (41)$$

We now apply this Proposition to the expression in Eq. (40). Let  $\mathbf{U} = \mathbf{L}\mathbf{D}_2^{1/2}$  and  $\mathbf{U}^T \mathbf{X} \mathbf{U} = \mathbf{D}_2^{1/2} \mathbf{L}^T \mathbf{X} \mathbf{L} \mathbf{D}_2^{1/2}$  for any matrix  $\mathbf{X}$ , so that  $\mathbf{H}$  has the form given in Eq. (35). In addition, let  $\mathbf{K} = \tilde{A}\mathbf{W}$ , and assume that  $\Psi(\tilde{A}\mathbf{W})$ ,  $\Psi(\tilde{A}\Theta\mathbf{W}) < 1$ . Then Prop. 5.1 gives

$$\begin{aligned} \langle \tilde{\mathbf{C}}^\infty \rangle_B / \tilde{C}^0 &= \mathbf{L}^T (\mathbf{I} - \tilde{A}\mathbf{W})^{-1} (\mathbf{I} - \tilde{A}\mathbf{W}^T)^{-1} \mathbf{L} \\ &= \left( \mathbf{I} - \sum_{n=1}^{\infty} \tilde{A}^n \mathbf{L}^T \mathbf{W}_n \mathbf{L} \mathbf{D}_2 \right)^{-1} \left( \mathbf{D}_2^{-1} + \sum_{n,m=1}^{\infty} \tilde{A}^{n+m} \mathbf{L}^T \mathbf{W}_n \Theta \mathbf{W}_m^T \mathbf{L} \right) \\ &\quad \cdot \left( \mathbf{I} - \sum_{m=1}^{\infty} \tilde{A}^m \mathbf{D}_2 \mathbf{L}^T \mathbf{W}_m^T \mathbf{L} \right)^{-1}, \end{aligned} \quad (42)$$

where

$$\mathbf{W}_n = \underbrace{\mathbf{W} \Theta \mathbf{W} \cdots \Theta \mathbf{W}}_{n \text{ factors of } \mathbf{W}}.$$

As in the single population case, we can discard all terms in Eq. (42) which do not correspond to second order motif frequencies. This means that in the first and third set of brackets in Eq. (42), we discard any terms containing  $\mathbf{W}_n$  with  $n \geq 3$ , and for the middle set of brackets, we discard terms containing  $\mathbf{W}_n \Theta \mathbf{W}_m^T$  for  $n + m \geq 3$ . This gives

$$\langle \tilde{\mathbf{C}}^\infty \rangle_B / \tilde{C}^0 \approx \left( \mathbf{I} - \tilde{A} \mathbf{M} \mathbf{D}_2 - \tilde{A}^2 \mathbf{Q}_{\text{ch}} \mathbf{D}_2 \right)^{-1} \left( \mathbf{D}_2^{-1} + \tilde{A}^2 \mathbf{Q}_{\text{div}} \right) \left( \mathbf{I} - \tilde{A} \mathbf{D}_2 \mathbf{M}^T - \tilde{A}^2 \mathbf{D}_2 \mathbf{Q}_{\text{ch}}^T \right)^{-1}. \quad (43)$$

Figure 12 illustrates that this approximation is a great improvement over that given by truncating the expansion at second order (compare with Fig. 11). Again, we note that our approximation requires only knowledge of the overall connection probabilities among excitatory and inhibitory cells, and the frequency of second-order motifs.

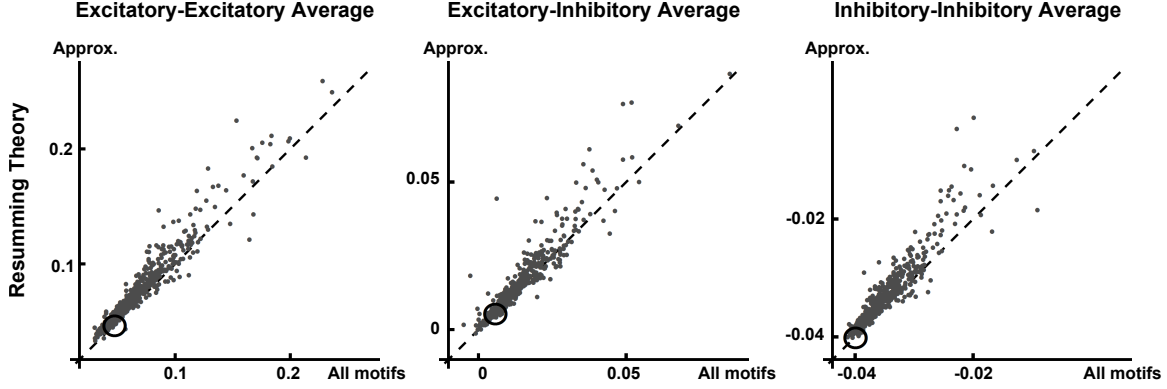


Figure 12: Predicting block-wise average correlations from resumming theory. The horizontal-axis is the average correlation (in a certain block) from original linear response expression of covariance matrix (Eq. 7); the vertical axis the quantity from resumming theory (Eq. (43)). The diagonal line  $y = x$  is plotted for reference. The spectral radius of  $\tilde{A}\mathbf{W}$  under Erdős-Rényi assumption is 0.33 (see Sec. 2.4). Other network parameters are the same as in Fig. 10. Here  $R^2 = 0.93, 0.88, 0.87$  respectively for the three panels. The open circle indicates the level of correlation expected from an Erdős-Rényi network with the same overall connection probability and strength.

Expanding the inverses in Eq. (43) in a power series, we can again obtain an approximation to block average correlations to linear order in  $\mathbf{Q}_{\text{ch}}$ , and  $\mathbf{Q}_{\text{div}}$ ,

$$\begin{aligned} \langle \tilde{\mathbf{C}}^\infty \rangle_B / \tilde{\mathbf{C}}^0 \approx & \left( \mathbf{I} - \tilde{\mathbf{A}}\mathbf{M}\mathbf{D}_2 \right)^{-1} \left[ \mathbf{D}_2^{-1} + \tilde{\mathbf{A}}^2 \mathbf{Q}_{\text{ch}} \left( \mathbf{I} - \tilde{\mathbf{A}}\mathbf{D}_2 \mathbf{M} \right)^{-1} + \left( \mathbf{I} - \tilde{\mathbf{A}}\mathbf{M}^T \mathbf{D}_2 \right)^{-1} \tilde{\mathbf{A}}^2 \mathbf{Q}_{\text{ch}}^T \right. \\ & \left. + \tilde{\mathbf{A}}^2 \mathbf{Q}_{\text{div}} \right] \left( \mathbf{I} - \tilde{\mathbf{A}}\mathbf{D}_2 \mathbf{M}^T \right)^{-1}. \end{aligned} \quad (44)$$

As in the single population case, each entry of the  $2 \times 2$  matrix on the right hand side of Eq. (44) gives an approximation to the block-averaged correlations expressed in terms of the scalars  $q_{\text{ch}}^{ZYX}, q_{\text{div}}^{XY,Z}$ , providing an analytical estimate to the regression coefficients plotted in Fig. 10. An example with uniform connection probability is given in Appendix G. We note that in general all sub-types of diverging and converging motifs affect each block-wise average correlation. This is not predicted by the second order truncation in Eq. (39). Somewhat counterintuitively,  $q_{\text{div}}^{EI,E}$  and  $q_{\text{div}}^{EI,I}$  (index 3 and 4 in Fig. 4) can contribute negatively to  $\langle \tilde{\mathbf{C}}_{EE} \rangle$  as shown in Fig. 10.

As in the single population case, we can also retain contributions to mean correlation which can be expressed as (nonlinear) functions of first order motifs (connection probabilities) only. This follows from setting all  $\mathbf{Q}$  terms in Eq. (43) to zero, yielding the following approximation for mean correlation in the two-population analog of Erdős-Rényi networks:

$$\langle \tilde{\mathbf{C}}^\infty \rangle_B / \tilde{\mathbf{C}}^0 = \left( \mathbf{I} - \tilde{\mathbf{A}}\mathbf{M}\mathbf{D}_2 \right)^{-1} \mathbf{D}_2^{-1} \left( \mathbf{I} - \tilde{\mathbf{A}}\mathbf{D}_2 \mathbf{M}^T \right)^{-1}. \quad (45)$$



This predicted mean correlation for an Erdős-Rényi network is shown by an open dot in Fig. 12; the deviations from this value illustrate that motif structures can both significantly increase and decrease average network-correlations (see also Fig. 2 and 9).

## 6 Heterogeneous networks

For simplicity in the previous sections we assumed a homogeneous network of neurons composed of cells with identical firing rates, power spectra and response properties. As a result, the diagonal matrices  $\tilde{\mathbf{C}}^0$  and  $\tilde{\mathbf{A}}$  in our key expression for correlations, Eq. (6), were scalar matrices,  $\tilde{C}^0 I$  and  $\tilde{A} I$ , and could be factored – leaving the connectivity structure of the network to determine the correlation value. In particular, this structure impacts correlations via the matrix products of adjacency matrices in Eq. (16).

Biological neural networks are heterogeneous. Even if the neurons are identical before they are coupled, any heterogeneities in the coupling structure will lead to different neurons firing with different rates and hence power spectra (cf. Zhao et al. [2011], which we return to in the Discussion). Moreover, as a consequence, they will also have different levels of responsivity. In sum, the matrices  $\tilde{\mathbf{C}}^0$  and  $\tilde{\mathbf{A}}$  will not have identical entries on the diagonal.

Consequently, the equivalent of the expansion Eq. (16) takes the form

$$\tilde{\mathbf{C}}^\infty = \sum_{i,j=0}^{\infty} (\tilde{\mathbf{A}}\mathbf{W})^i \tilde{\mathbf{C}}^0 (\mathbf{W}^T \tilde{\mathbf{A}}^*)^j. \quad (46)$$

As a result, a purely graph theoretic interpretation of the terms in the expansion – that is, one based on the connectivity *alone* – is no longer possible. Recall that as in Eqns. (9-11), motif frequencies are associated with terms like  $\mathbf{W}^i (\mathbf{W}^T)^j$ . Here the powers  $(\tilde{\mathbf{A}}\mathbf{W})^i$  are of a “weighted” connection matrix (with weights corresponding to the responsivity of different cells). Moreover,  $\tilde{\mathbf{C}}^0$  is no longer a scalar matrix and in general does not commute with  $\mathbf{W}$ , introducing additional complications that we address below.

In this section we discuss how to extend our results to the case of heterogeneous neural populations.

### 6.1 Performance of the homogeneous approximation

A first attempt at coping with heterogeneity is to hope that it is unimportant, and to apply the homogeneous theory of Eq. (16) naively. To do this, we need to choose approximate (scalar) values for the power spectrum  $\tilde{C}^0$  and responsivity  $\tilde{A}$  that we will apply to all of the cells, by plugging into the homogeneous network formula given by Eq. (16). We choose the  $\tilde{C}^0$  as the unadjusted power spectrum (due to the normalization (2), the actual value of  $\tilde{C}^0$  is not important), rather than using one homogeneous value for all cells. For  $\tilde{A}$ , we use the geometric mean of the  $\tilde{A}_i$ , which are found from self consistent equations similar to Eq. (5) (see [Trousdale et al., 2012] for details). The results of this naive application of the homogeneous theory are shown in panel **A** of Fig. 13. Although general trends are captured, this approach does not give an accurate approximation.

We note that here coupling is relatively strong (Erdős-Rényi formula for spectral radius giving 0.4, see Sec. 2.4). With weaker coupling (spectral radius=0.3), and hence less network-driven heterogeneity, the homogeneity assumption provides improved approximations in the heterogeneous case ( $R^2$  measure 0.78, 0.67, 0.67 for EE, EI and II average correlation respectively, data not shown). To quantify the heterogeneity of cellular dynamics across the networks, we compute the coefficient of variation  $CV$  of  $\tilde{A}_i$  and  $\tilde{C}_i^0$  averaged over network samples ( $CV = \langle \text{standard dev.} / \text{mean} \rangle$ ). For spectral radius=0.4, the CV is 0.35 and 0.38 for  $\tilde{A}_i$  and  $\tilde{C}_i^0$ , respectively, while at spectral radius=0.3, the CV is 0.23 and 0.26 for  $\tilde{A}_i$  and  $\tilde{C}_i^0$ . At such levels of heterogeneity, we clearly need a more systematic approach, and we develop this next.

## 6.2 Heterogeneous theory

We are prevented from applying Proposition 4.1 to Eq. (46) in the heterogeneous case, due to the presence of the factor  $\tilde{\mathbf{C}}^0$  in the middle of the terms on the right hand side. To deal with this difficulty we use the substitution  $\tilde{\mathbf{C}}^0 = (\tilde{\mathbf{C}}^0)^{1/2}(\tilde{\mathbf{C}}^0)^{1/2}$ , which is possible because the power spectrum is non-nonnegative. We can then rewrite Eq. (6) as

$$\begin{aligned} \tilde{\mathbf{C}}(0) &\approx (\tilde{\mathbf{C}}^0)^{1/2}(\tilde{\mathbf{C}}^0)^{-1/2}(\mathbf{I} - \tilde{\mathbf{A}}\mathbf{W})^{-1}(\tilde{\mathbf{C}}^0)^{1/2}(\tilde{\mathbf{C}}^0)^{1/2}(\mathbf{I} - \mathbf{W}^T\tilde{\mathbf{A}})^{-1}(\tilde{\mathbf{C}}^0)^{-1/2}(\tilde{\mathbf{C}}^0)^{1/2} \\ &= (\tilde{\mathbf{C}}^0)^{1/2} \left( \mathbf{I} - (\tilde{\mathbf{C}}^0)^{-1/2}\tilde{\mathbf{A}}\mathbf{W}(\tilde{\mathbf{C}}^0)^{1/2} \right)^{-1} \left( \mathbf{I} - (\tilde{\mathbf{C}}^0)^{1/2}\mathbf{W}^T\tilde{\mathbf{A}}(\tilde{\mathbf{C}}^0)^{-1/2} \right)^{-1} (\tilde{\mathbf{C}}^0)^{1/2}, \end{aligned} \quad (47)$$

where we are again evaluating all quantities at  $\omega = 0$ , so that  $\tilde{\mathbf{F}} = \mathbf{I}$ .

Let  $\tilde{A}_0$  be the geometric mean of  $\tilde{A}_i$ , which we choose to normalize responsivity so that weighted quantities will have the same units (see below). We can then define an effective or “functional” connection matrix, which has the same units as  $\mathbf{W}$ :

$$\hat{\mathbf{W}} = (\tilde{\mathbf{C}}^0)^{-1/2}\tilde{\mathbf{A}}\mathbf{W}(\tilde{\mathbf{C}}^0)^{1/2}/\tilde{A}_0, \quad (48)$$

so that Eq. (47) becomes

$$(\tilde{\mathbf{C}}^0)^{1/2}(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}})^{-1}(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}}^T)^{-1}(\tilde{\mathbf{C}}^0)^{1/2}. \quad (49)$$

This expression will be much easier to study, as the diagonal matrix  $\tilde{\mathbf{C}}^0$  no longer appears in the middle. We can thus expand the two terms,  $(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}})^{-1}$  and  $(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}}^T)^{-1}$  to obtain an expression that has a form analogous to that of Eq. (16).

The only difference in the present case of a heterogeneous network is that the definition of motif frequency and connection probability will involve weighted averages. For example, the entries of  $\hat{\mathbf{W}}$  are scaled version of entries of the connection matrix  $\mathbf{W}$ :

$$\hat{W}_{ij} = \frac{\tilde{A}_i}{\tilde{A}_0} \frac{\sqrt{\tilde{S}_j}}{\sqrt{\tilde{S}_i}} W_{ij}.$$

As an example, a diverging motif  $i \leftarrow k \rightarrow j$  has a weighted contribution of the form

$$\frac{\tilde{A}_i \tilde{A}_j}{\tilde{A}_0^2} \frac{\tilde{C}_k^0}{\sqrt{\tilde{C}_i^0 \tilde{C}_j^0}} W_{ik} W_{jk}. \quad (50)$$

The ratio  $(\tilde{A}_i \tilde{A}_j)/\tilde{A}_0^2$  in Eq. (50) quantifies the relative responsivity of the recipient cells. Hence, a particular diverging motif will be weighted more strongly (and provide a greater contribution to average correlation) if the recipient cells are more responsive to inputs. Similarly,  $\tilde{C}_i^0, \tilde{C}_j^0, \tilde{C}_k^0$  corresponds to the variance of spike counts in long time windows for the uncoupled cells: the weight is determined by the variance of the projecting cell divided by the geometric mean of the variance of the recipient cells. These observations agree well with intuition: more responsive cells will be more strongly correlated by a common input, and “source cells” with larger total variance ( $\tilde{C}_k^0$ ) will lead to a diverging motif with larger impact.

The motif frequencies  $q_{\text{div}}$  and  $q_{\text{ch}}$ , are defined by Eq. (18) and (20) upon substituting  $\mathbf{W}^0$  with  $\hat{\mathbf{W}}^0 = \hat{\mathbf{W}}/w$  in the single population case. In the case of two populations, the matrices  $\mathbf{Q}_{\text{div}}$  and  $\mathbf{Q}_{\text{ch}}$  are defined by Eq. (36) and (38), with  $\mathbf{W}$  replaced by  $\hat{\mathbf{W}}$ . These weighted motif frequencies could be estimated experimentally in an active network, via recordings from neurons known to participate in three cell motifs. The advantage of our resumming theory remains that everything needed is based only on the statistics of a relatively small number of cell motifs, rather than higher order information about the connectivity graph.

Due to the normalization in Eq. (2), the two outside diagonal matrix factors in Eq. (49) cancel in the definition of  $\boldsymbol{\rho}$ , and we can use the following matrix to calculate correlation coefficients (noting that we continue to approximate  $\tilde{\mathbf{C}}_{ii}$  by  $\tilde{C}_i^0$  when performing the normalization in the denominator of Eq. (2)):

$$(\mathbf{I} - \tilde{A}_0 \hat{\mathbf{W}})^{-1} (\mathbf{I} - \tilde{A}_0 \hat{\mathbf{W}}^T)^{-1}. \quad (51)$$

Eq. (51) is the heterogeneous analog of the expression for the correlation  $\tilde{\mathbf{C}}^\infty/\tilde{\mathbf{C}}^0$  studied in Sections 4 and 5 (see Eq. (7)), including diagonal contributions. Applying the motif resumming theory exactly as in these sections yields corresponding approximations of mean correlation for these systems. In particular, for the two population case, the mean correlation  $\rho_{EE}^{\text{avg}}, \rho_{EI}^{\text{avg}}, \rho_{II}^{\text{avg}}$  will be estimated based on Eq. (43) with the substitution of weighted motif counts  $\mathbf{Q}_{\text{div}}$  and  $\mathbf{Q}_{\text{ch}}$  (note that one must still adjust the approximation to account for diagonal terms — see Appendix A). In panel **B** of Fig. 13, we plot this prediction of  $\rho_{EE}^{\text{avg}}, \rho_{EI}^{\text{avg}}, \rho_{II}^{\text{avg}}$  compared with the full expression for correlation in the heterogeneous case (via Eqns. (2) and (6)). Accounting for dynamical heterogeneity across the network by defining *weighted* second order motifs provides a reasonably accurate prediction of average correlation.

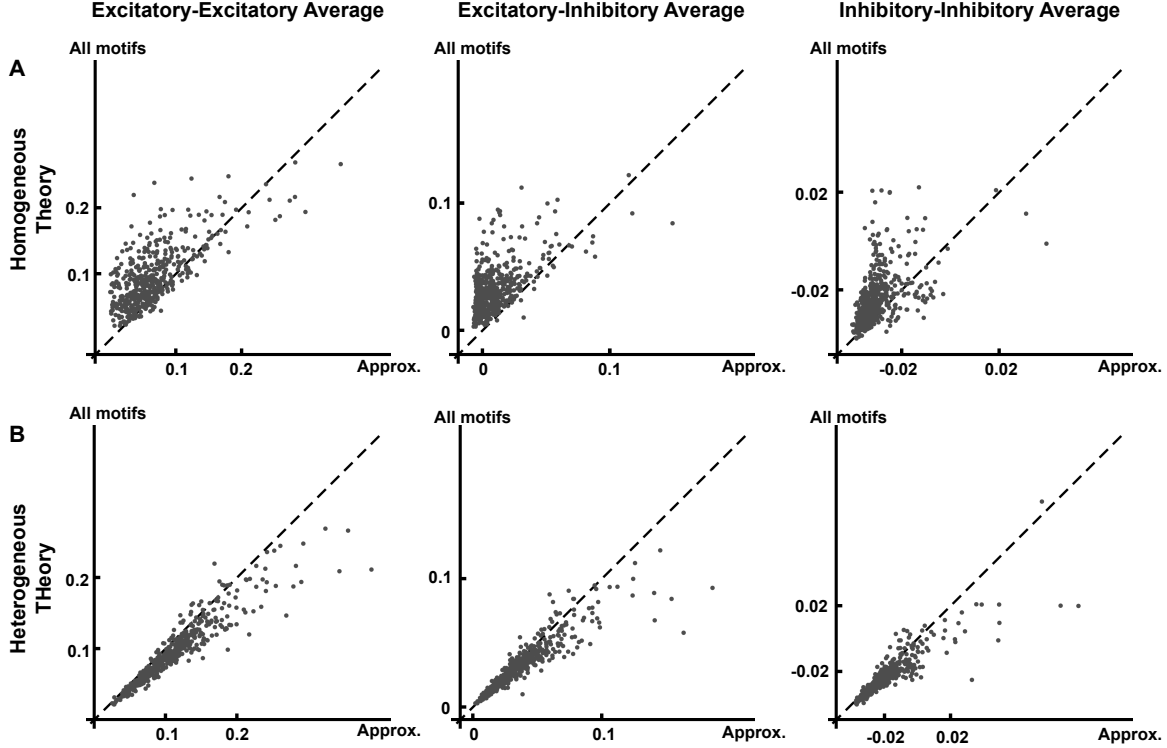


Figure 13: **A** Mean correlation from homogeneous resumming theory (horizontal axis) comparing with that from full linear response theory (Eq. (7)) (vertical axis). **B** Mean correlation from heterogeneous resumming theory (horizontal-axis) comparing with that from full linear response theory (vertical-axis). The diagonal line  $y = x$  is plotted for reference. The spectral radius of  $\tilde{A}\mathbf{W}$  under Erdős-Rényi assumption is 0.40 (see Sec. 2.4). Other network parameters are given in the caption of Fig. 10. The coefficients of determination,  $R^2$ , are 0.56, 0.44, 0.35 in panel **A** and 0.88, 0.81, 0.73 for panel **B**.

## 7 Comparisons with IF simulations

In the preceding sections, we have shown how network-wide correlation can be approximated based on coupling probability together with the frequency of second-order motifs. In assessing the accuracy of this motif-based theory, we have compared the predictions of the theory with the value of correlation given Eq. (6), which is exact in the sense that it precisely includes the contribution of motifs at all orders and therefore gives an exact description of linearly interacting point process models [Pernice et al., 2011, Hawkes, 1971a].

When using our theory to predict the impact of motifs on mean correlation for networks of IF neurons, we are making an additional approximation in describing integrate-and-fire dynamics with linear response theory [Trousdale et al., 2012, Ostojic et al., 2009, Lindner et al., 2005]. We now directly test the performance of our motif-based predictions in IF networks – thus probing how errors at each of the two layers of approximation combine. Specifically, in Fig. 14 we compare the block-wise mean correlations from IF simulation and

the predictions obtained using Eq. (43). These simulations are for the same networks used in Fig. 3. We find that our theory gives a good quantitative match with LIF simulations, despite the multiple approximations involved. Thus, our theory predicts trends in the impact of motif frequencies on network-wide correlation.

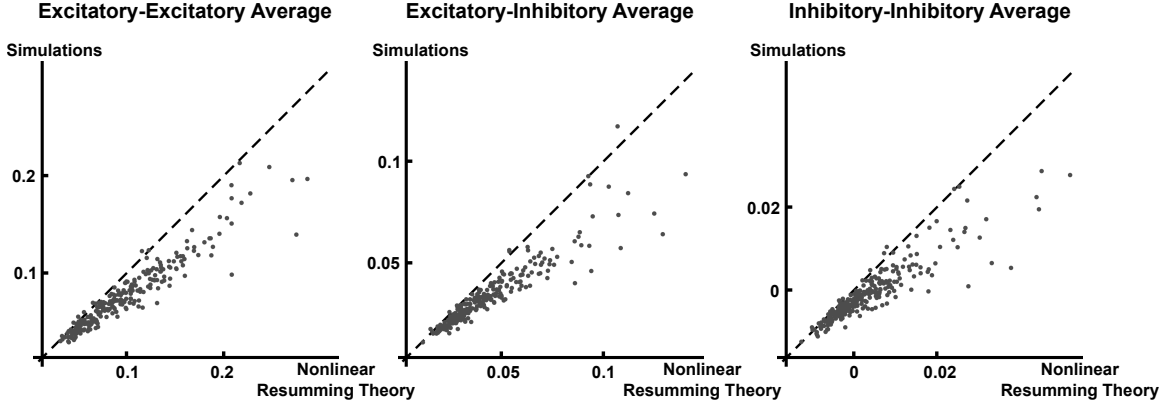


Figure 14: Average correlation from integrate and fire neuron simulations compared with the predictions of Eq. (43). The vertical axis is the mean correlation coefficient calculated from simulations of the same 265 excitatory-inhibitory networks studied in Fig. 3. The horizontal axis is the prediction for mean correlation from the resummation theory based on empirical connection probability and second order motif frequencies. The diagonal line  $y = x$  is plotted for reference. The spectral radius of  $\tilde{A}\mathbf{W}$  under the Erdős-Rényi assumption is 0.33 (see Sec. 2.4). Coefficients of determination  $R^2$  are 0.91, 0.87, 0.79 respectively for the three panels.

## 8 Discussion

### Summary: Predicting network-wide correlation from three cell motifs

We studied the impact of the graphical structure of neural networks – characterized by connectivity motifs – on their dynamical structure, characterized by correlated activity between cells. As shown in Fig. 2, varying the frequency of such motifs can strongly impact correlation, over and above the overall level of connectivity in a network. Following Zhao et al. [2011], we focus on the three types of motifs that involve two connections each: the diverging, converging and chain motifs.

We chose a standard spiking neuron model, the integrate-and-fire (IF) neuron, in constructing our recurrent networks (Sec. 2). For IF neurons [Lindner et al., 2005, Trousdale et al., 2012], LNP neurons [Beck et al., 2011], and other neuron models such as linearly interacting point process model [Hawkes, 1971a,b], one can apply linear response approximations (or, in the case of Hawkes models, solve exact equations – see “Neuron models” below) to get

an explicit expression for the pairwise correlations depending explicitly on the connectivity matrix (see Eq. (6)).

We expand this expression in a series, where each term has clear correspondence to certain graphical structures Pernice et al. [2011], Rangan [2009a]. In particular, second order terms correspond to second order motifs. Importantly, we show that contributions of higher order terms can be estimated using the frequencies of second order motifs along with the connection probability.

For systems with correlations well-approximated by Eq. (6) and assuming homogeneous cellular dynamics, we find that the frequency of converging motifs, over and above that expected in Erdős-Rényi networks, will have no effect on mean correlation in systems (if the connection probability and other motif frequencies are fixed). Meanwhile diverging and chain motifs will contribute positively. In networks of excitatory and inhibitory neurons, the three types of motifs are subdivided according to the type of the constituent neurons. However, average correlations between cells of a given type are still given in terms of diverging and chain motifs (see Fig. 3).

We first analyzed networks in which we made a strong homogeneity assumption on the dynamical properties of the uncoupled neurons. The resumming theory we develop approximates the contributions of higher order motifs in terms of the frequency of second order motifs. In Sec. 6, we extended our theory to heterogeneous networks. In such cases, the contribution of one instance of a certain motif to the total motif frequency will be additionally weighted by the relative responsiveness of the neurons composing the motif which *receive* input, as well as the baseline variance of the cells. Overall, our results can be regarded as a general estimator of mean correlation given motif statistics up to second order.

To test our theory numerically, we generated random networks of IF neurons with fixed statistical (i.e., expected) connection probability, but different second order connectivity statistics (see Sec. 3). Simulations show that the theory provides an accurate description of network correlations (see Fig. 14), despite the additional error introduced by the linear response approximation of activity (Eq. (4)). We also compared the resumming theory to the direct evaluation of the linear response theory (Eq. (6)) which takes account of the full graphical structure. The close match between the two (Fig. 9, Fig. 12) shows that second order motifs capture much of the dependence of mean correlation on network connectivity. Moreover, Figs. 2, 9, 12 demonstrate that variations in the frequency of second order motifs produce changes in network correlation that dominate those expected from the small variation in empirical (i.e., realized) connection probability from one network to the next.

Beyond the resumming theory, we also considered two other ways of simplifying the expansion of Eq. (6). The first was a truncation in connection strength at second-order (in powers of  $\tilde{A}$  and  $\mathbf{W}$ ; Secs. 4.2, 5.2). This eliminates all contributions due to motifs of length two or less. Except in very weakly connected networks, this is a poor approximation: Although the contributions of higher order motifs decay exponentially in interaction strength  $\tilde{A}w$ , their number also grows exponentially with  $Np$ . Thus, motifs of all orders could, in principle, contribute substantially to average correlations. Our second truncation was the Erdős-Rényi approximation of mean correlation. This yielded predictions for mean correla-

tion that included contributions from motifs of all orders. These predictions therefore depend only on connection probability (See Eqns. (32, 45)), and should be valid when there is very little structure in the connectivity graph compared to an idealized Erdős-Rényi network.

## Neuron models

Our motif-based theory can be applied to a variety of different neuron models. The starting point of our theory, the expression for pairwise cross-spectra given by Eq. (6), arises in a number of settings.

The main tool in our approach is linear response theory. This connects our methods to IF models (see Sec. 2.3). Importantly, Eq. (6) also arises as an *exact* expression for linearly interacting point process, or *Hawkes* models [Hawkes, 1971b, Pernice et al., 2011, 2012]. In this case,

$$\tilde{\mathbf{C}}(\omega) = (\mathbf{I} - \tilde{A}(\omega)\mathbf{W})^{-1}\mathbf{Y}(\mathbf{I} - \tilde{A}(\omega)\mathbf{W}^T)^{-1},$$

where  $\tilde{A}(\omega)$  is the Fourier transform of the interaction filters and  $\mathbf{Y}$  is a diagonal matrix of firing rates. Therefore our analysis can be directly and exactly applied to this setting. We note that the Hawkes process is a linear-Poisson (LP) model. Such models are commonly applied in theoretical neuroscience [Dayan and Abbot, 2001].

## Relationship to other studies on motifs and networks correlation

Our methodology is very similar to the previous study of Zhao et al. [2011]. They considered the impact of diverging, converging and chain motifs (along with the reciprocal connection motif) on the ability of an excitatory recurrent network to stay synchronized. They found that prevalence of the converging motif decreases synchrony, prevalence of the chain motif increases synchrony, and that the diverging motif has no significant effect on synchrony. The difference between their results and ours can be understood from the different dynamical regimes considered. Zhao et al. [2011] considered perturbations from two extreme cases: perfect synchrony, and evenly distributed asynchronous oscillators. In contrast, we studied the asynchronous regime but allowed the activity of the cells to be correlated. Hence, our methods also differ: we use a linear response approach valid for strong internal noise, and weak to intermediate coupling, while Zhao et al. [2011] performed a linear stability analysis of coupled oscillator equations.

Many other studies have also examined the relationship between graphical features and spike correlations in networks. Roxin [2011] studied how spike-time correlation changes when one interpolates the degree distribution of a network from the binomial distribution (corresponding to Erdős-Rényi networks) to a truncated power law distribution. He found that increasing the variance of the out-degree distribution will increase the cross-correlation, which is consistent with our results for the diverging motif (see Eq. (12)).

Pernice et al. [2011] and Trousdale et al. [2012] studied the influence of connection structures on pairwise correlation for a number of network classes. Pernice et al. [2011] used the Hawkes neuron model, which as noted above, leads to very similar expressions for correlation

as those studied here and by Trousdale et al. [2012]. Both approaches relate pairwise correlations to certain graphical structures of increasing order (i.e., motif size). In particular, Pernice et al. [2011] obtained an expression for average network correlation in terms of the mean input and common input (Eq. (22) in [Pernice et al., 2011]) for regular networks with uniform connection probability, while Trousdale et al. [2012] (Eq. (25) in Trousdale et al. [2012]) considered correlations in networks where only in-degree was fixed. Both are special cases of our resumming theory (Note that according to Eq. (12) and (13), a fixed in (resp. out) degree is equivalent to  $q_{\text{con}} = 0$  (resp.  $q_{\text{div}} = 0$ ) and  $q_{\text{ch}} = 0$ ).

A major contribution of the present study is to show that effects of “higher order” graphical interactions (i.e., motifs including more than three cells) can be approximated in terms of the frequency of second order motifs and the overall connection probability. This allows a systematic treatment of network-averaged correlation for a broader range of network connectivities.

## Limitations

When applied to integrate and fire neuron networks, our analysis relies on the validity of the linear response approximation. In Section 7, we demonstrated this validity for a particular firing regime, for the class of random networks studied here. For more on this issue, see [Trousdale et al., 2012] and [Ostojic and Brunel, 2011]. We note that one avoids this issue entirely when considering Hawkes processes (see above).

We also assumed that the spectral radius of the total coupling matrix is less than one, in order to expand Eq. (6) into a series, for which each term can be attributed to a different motif. From this point, our methods rely on our ability to predict the impact of higher order motifs on network correlation based on the frequency of second order motifs. We demonstrated that our resumming theory can successfully make this prediction for classes of networks generated in two ways: via two-sided power law distributions, and via the SONENT method (see Sec. 3.2). However, for certain connectivity matrices, our resumming theory can produce large errors. An example pointed out to us by Chris Hoffman [personal communication] is  $\mathbf{W}^0$  containing a  $\sqrt{p}N \times \sqrt{p}N$  block of 1 entries and but taking value 0 everywhere else; here,  $p$  is the overall connection probability. Note that this matrix corresponds to a graph with one fully connected group and another fully isolated group. Such disconnected architecture and strong inhomogeneity may be features that produce large spectral radii  $\Psi(\mathbf{W}\Theta)$ , and hence large errors when applying our theory (see also Appendix H; and for a study on the dynamics of inhomogeneous clustered networks, [Watts and Strogatz, 1998]).

Three other factors limit the generality of our results. First, in order to separate the effect of network structure from that of cellular dynamics we initially assumed homogeneous firing rates and identical neurons. For many real neural networks, this may be a poor approximation. Thus, in our analysis of heterogeneous networks, we calculate weighted motif strengths, *assuming* “*a priori*” knowledge of the (heterogeneous) cellular firing rates and responsivities. A full theory would rather begin by *predicting* this heterogeneity based on network properties. Intriguingly, [Zhao et al., 2011] show that certain motif frequencies can be a powerful source of such heterogeneity, a potentially important fact that we neglect.



Second, for simplicity we only considered long-time window spike count correlation. However, our analysis can be easily generalized to study correlation at any time scale, and for any  $\omega$  in Eq. (6). A third, and final limitation of our analysis is that we only studied pairwise correlation coefficients. This may not adequately describe the network dynamics or reveal the full importance of certain motifs.

## Extensions and connections with neurobiology and computation

Intriguingly, experiments have shown that motif statistics in actual neural networks deviate from the Erdős-Rényi network model, opening a possible role of multicellular motifs. For example, Song et al. [2005] found such deviations in connectivity between excitatory cells in visual cortex. Sporns and Kötter [2004] also suggest that increased frequencies of certain motifs may be a common feature of neural circuits. Moreover, [Haeusler et al., 2009] studied the different motif statistics from two experimental data of connectivity structures of laminae in cortex, showing deviations from expectations under the Erdős-Rényi model without the laminar structure (implying as a possible origin of non-trivial motif frequencies). Our study can be applied to suggest a possible consequence of these experimental findings for levels of spike-time correlation (in the latter case generalizing it to apply to multi-group networks corresponding to the different laminae).

[Renart et al., 2010] studied pairwise correlations in balanced excitatory-inhibitory Erdős-Rényi networks, and found that balanced neuronal networks will have average correlations that decrease rapidly to zero with network size. Our results suggest that, with increased propensity of certain second order motifs, EI networks can potentially produce significantly larger correlations value (see Fig. 2). However, to apply our findings to the case considered by [Renart et al., 2010], we would need to make several extensions, including allowing full heterogeneity arising from recurrent connections and checking that linear response theory works in their strongly coupled, balanced dynamical states. In our setting, “balanced” can be defined as that the total mean input received for each cell is 0, exactly or on average across realizations of the network [Rajan and Abbott, 2006]. This needs to be compared with the dynamic balance concept in [Renart et al., 2010].

For future studies, we also hope to develop a theory that predicts not just the mean correlation strength across a network, but its variance across cell pairs. This variance has been studied using theoretical and experimental approaches [Renart et al., 2010, Ecker et al., 2011], and it would be interesting to describe how it depends on connection motifs. We will also try to predict the heterogeneity in cellular dynamics caused by motif frequencies, as referred to above – incorporating, for example, the result emphasized in [Zhao et al., 2011] that variability in input to different neurons depends on converging motifs (Eq. (13)). Finally, we note the important connections between correlations and coding efficiency [Zohary et al., 1994, Abbott and Dayan, 1999, Salinas and Sejnowski, 2000, Josić et al., 2009, Schneidman et al., 2006] (see also Introduction). The recent work of [Beck et al., 2011] set up a direct connection between the graphical structure of networks and their coding efficiency, using a similar linear response calculation of covariance matrix for linear nonlinear Poisson (LNP) neurons. The present results could be used with the approach of Beck et al. [2011] to further

link the statistics of motifs to properties of signal transmission in neural networks.

## A Approximating $\rho^{\text{avg}}$ from $\langle \tilde{\mathbf{C}}^\infty \rangle$

The average covariance across the network,  $\langle \tilde{\mathbf{C}}^\infty \rangle$  can be used to approximate  $\rho^{\text{avg}}$ . Here we describe two such approximations. First, when the uncoupled neurons have equal firing rates, and the perturbation from recurrent coupling is weak, the diagonal terms  $\tilde{\mathbf{C}}_{ii}^\infty$  will be close to the unperturbed values,  $\tilde{C}^0$ . In this case, subtracting diagonal terms from the average, we have that

$$\rho^{\text{avg}} \approx \left( \frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0} - \frac{1}{N} \right) \cdot \frac{N}{N-1}. \quad (52)$$

In a second, more accurate approximation, we assume permutation symmetry between neurons within one population. Also, in our networks, self-connections are allowed and occur with the same probability as other connections. These will lead to identical (marginal) distributions for each entry in the covariance matrix  $\tilde{\mathbf{C}}^\infty$ , excepting the diagonal entries which are shifted by a constant of  $\tilde{C}^0$  due to each neuron's own unperturbed variance (this corresponds to the term proportional to  $\mathbf{I}$  if one were to expand Eq. (7) as a series — see Eq. (16)). This suggests that  $\tilde{\mathbf{C}}^\infty$  has the form

$$\tilde{\mathbf{C}}^\infty \approx \tilde{C}_0 \mathbf{I} + c \mathbf{1}_{NN},$$

where  $\mathbf{1}_{NN}$  is the  $N \times N$  matrix of all ones, and  $c$  is a constant. If this holds, then, using diagonal entries of this matrix as normalization, we obtain

$$\rho^{\text{avg}} \approx \frac{\langle \tilde{\mathbf{C}}^\infty \rangle - \tilde{C}_0/N}{\tilde{C}_0 + \langle \tilde{\mathbf{C}}^\infty \rangle - \tilde{C}_0/N}. \quad (53)$$

The two approximations given in Eqns. (52,53) are approximately equal for small correlations. We will use Eq. (52) to exhibit the linear dependence between mean correlation coefficient and motif frequency (such as linear weights in Fig. 3 Part B) and Eq. (53) for quantitative predications (all numerical plots).

For networks consisting of an excitatory and inhibitory population, we have analogs of Eqns. (52,53) for each of the 4 population blocks of the covariance matrix,

$$\begin{aligned} \rho_{XX}^{\text{avg}} &\approx \left( \frac{\langle \tilde{\mathbf{C}}_{XX}^\infty \rangle}{\tilde{C}^0} - \frac{1}{N_X} \right) \cdot \frac{N_X}{N_X - 1}, \\ \rho_{XY}^{\text{avg}} &\approx \langle \tilde{\mathbf{C}}_{XY}(0) \rangle / \tilde{C}(0), \end{aligned} \quad (54)$$

and

$$\begin{aligned} \rho_{XX}^{\text{avg}} &\approx \frac{\langle \tilde{\mathbf{C}}_{XX} \rangle - \tilde{C}_0/N_X}{\tilde{C}_0 + \langle \tilde{\mathbf{C}}_{XX} \rangle - \tilde{C}_0/N_X}, \\ \rho_{XY}^{\text{avg}} &\approx \frac{\langle \tilde{\mathbf{C}}_{XY} \rangle}{\sqrt{(\tilde{C}_0 + \langle \tilde{\mathbf{C}}_{XX} \rangle - \tilde{C}_0/N_X)(\tilde{C}_0 + \langle \tilde{\mathbf{C}}_{YY} \rangle - \tilde{C}_0/N_Y)}}, \end{aligned} \quad (55)$$

where  $X \neq Y \in \{E, I\}$ .

## B Proof of bound on $q_{\text{div}}, q_{\text{con}}, q_{\text{ch}}$ for one population

We here prove the inequalities in Eq. (15). Noting that we may write

$$q_{\text{div}} = \mathbf{E}_e[\mathbf{W}_{i,k}^0 \mathbf{W}_{j,k}^0] - \mathbf{E}_e[\mathbf{W}_{i,k}^0] \mathbf{E}_e[\mathbf{W}_{j,k}^0],$$

(with similar expressions holding for  $q_{\text{con}}, q_{\text{ch}}$ ) it is sufficient to show  $\mathbf{E}_e[\mathbf{W}_{i,k}^0 \mathbf{W}_{j,k}^0] \leq p$ . Note

$$\mathbf{H} \mathbf{W}^0 (\mathbf{W}^0)^T \mathbf{H} = (N^2 \mathbf{E}_e[\mathbf{W}_{i,k}^0 \mathbf{W}_{j,k}^0]) \mathbf{H}, \quad (56)$$

where  $\mathbf{H}$  is defined in Eq. (17). Let  $\|\cdot\|_F$  be the Frobenius norm, which is sub-multiplicative. We then have

$$\|\mathbf{H} \mathbf{W}^0 (\mathbf{W}^0)^T \mathbf{H}\|_F \leq \|\mathbf{H} \mathbf{W}^0\|_F \|(\mathbf{W}^0)^T \mathbf{H}\|_F \leq \|\mathbf{H}\|_F \|\mathbf{W}^0\|_F \|(\mathbf{W}^0)^T\|_F \|\mathbf{H}\|_F. \quad (57)$$

Since  $\|\mathbf{H}\|_F = 1$ ,  $\|\mathbf{W}^0\|_F = \sqrt{\sum_{i,j} (W_{i,j}^0)^2} = N\sqrt{p}$ , the above inequality, together with (56), gives  $\mathbf{E}_e[\mathbf{W}_{i,k}^0 \mathbf{W}_{j,k}^0] \leq p$ . With the fact that  $\mathbf{E}_e[\mathbf{W}_{i,k}^0] \mathbf{E}_e[\mathbf{W}_{j,k}^0] = p^2$ , we have the bound in Eq. (15).

In the second inequality of (57), we have used  $\|\mathbf{H} \mathbf{W}^0\|_F \leq \|\mathbf{H}\|_F \|\mathbf{W}^0\|_F$ . The necessary and sufficient condition for achieving equality is the equality condition in the following Cauchy inequalities for all  $i, j$

$$\left( \sum_k H_{i,k} W_{k,j}^0 \right)^2 \leq \left( \sum_k H_{i,k}^2 \right) \left( \sum_k (W_{k,j}^0)^2 \right).$$

Equality holds exactly when  $W_{k,j}^0 = W_{l,j}^0$  for all  $k, l$  – that is, each column of  $\mathbf{W}^0$  has same values (either all ones or all zeroes). The equality condition for the first inequality in Eq. (57) is identical. To generate an example graph that achieves equality for a specified overall connection  $p$ , we simply set a fraction  $p$  of the columns to be all ones, and set the remaining columns to zero.

Similarly, for converging motifs we can show  $\mathbf{E}_e[\mathbf{W}_{k,i}^0 \mathbf{W}_{k,j}^0] \leq p$  with the same equality condition as above, and for chain motifs,  $\mathbf{E}_e[\mathbf{W}_{i,k}^0 \mathbf{W}_{k,j}^0] \leq p$ . However, the equalities for  $q_{\text{div}}, q_{\text{con}}$  cannot hold simultaneously, and the equality for  $q_{\text{ch}}$  cannot be achieved.

## C Graph generation methods

Here we present more details on how we generated network samples with fixed connection probability, but different frequencies of second order motifs. We used two methods.

First, the degree distribution method consists of initially generating a sample of in and out degrees from a truncated power law distribution with density, following [Zhao et al., 2011],

$$f(d) = \begin{cases} C_1 d^{\gamma_1} & 0 \leq d \leq L_1 \\ C_2 d^{\gamma_2} & L_1 \leq d \leq L_2 \\ 0 & \text{otherwise} \end{cases}, \quad (58)$$

where  $d$  is the in or out degree (see also the configuration model [Roxin, 2011, Newman, 2003, Newman and Watts, 2001]). The two marginal distributions of in and out degree are then coupled using a Gaussian copula with correlation coefficient  $\rho$  to generate the in and out degree lists. The parameters  $\rho$ ,  $L_1/L_2$ ,  $L_2$ ,  $\gamma_1 > 0$ ,  $\gamma_2 < 0$  are independently and uniformly sampled for each network, separately for in and out degrees (their ranges are listed in Table 1).  $C_1$  and  $C_2$  are chosen so that  $f(d)$  is continuous at  $L_1$  and the mean of the degree distribution is normalized to  $Np$  (the same value for both in and out degrees), where  $p$  the fixed connection probability across network samples and  $N$  is the network size.

We then use the degree lists to calculate a probability for each possible connection from cell  $j$  to cell  $i$ , which is proportional to  $d_i^{\text{in}} d_j^{\text{out}}$ . All these  $N^2$  probabilities are scaled so that the resulting average for the total number of connections is the same as the quantity  $N^2 p_{\text{stat}}$ . Here  $p_{\text{stat}}$  is the target connection probability that we aim to achieve in these samples; we recall that the “empirical” connection probability achieved in a given graph is notated as  $p$ .

	$\rho$	$\gamma_1$	$\gamma_2$	$L_1/L_2$	$L_2/N$
min	-1	0.25	-2.25	0	0.7
max	1	2.25	-0.25	1	1

Table 1: Range of parameters of the truncated power law degree distribution.

For the excitatory-inhibitory case, we generate four degree lists  $d_E^{\text{in}}, d_E^{\text{out}}, d_I^{\text{in}}, d_I^{\text{out}}$ , again according to the marginal distributions (58). We therefore need a four dimensional Gaussian copula. Again, parameters for the power law distribution and the correlation coefficient matrix of the copula are randomly chosen. Using these degree list, we can then generate each of the four blocks (defined by cell type) of adjacency matrix in the same way as in the single population case. This method allows us to sample from the whole extent of motif parameters (3 in single population case, 20 in two population case).

For the single population networks studied in Sec. 4, we generated additional network samples via the SOnET method [Zhao et al., 2011]. The idea of the algorithm is similar to maximum entropy models. Given only the connection probability and second order motif frequency, we try to generate the most “random” network satisfying these moment constraints. However, instead of using the Gibbs distribution for the connections as in maximum entropy method, we use a dichotomized ( $N^2$  dimensional) Gaussian distribution. We then generate samples of excitatory only networks from the complete range of possible motif frequencies, as described in Eqns. (14)-(15), using the SOnET algorithm.

Network samples generated using both methods cover the range of motif frequency observed experimentally in cortical circuits [Song et al., 2005, Zhao et al., 2011] as shown by Table 2. Here, we list motif frequency values as  $(q_{\text{div}}, q_{\text{con}}, q_{\text{ch}})/(p(1-p))$  for excitatory only networks, and  $(q_{\text{div}}^{EE,E}, q_{\text{con}}^{EE,E}, q_{\text{ch}}^{EEE})/(p_{EE}(1-p_{EE}))$  in excitatory-inhibitory networks (since [Song et al., 2005] only recorded from excitatory neurons, see definition at Eq. (36)) .

		SONET: E only		degree: E only		degree: EI	
	experiment	min	max	min	max	min	max
diverging	0.033	0.001	0.913	0.015	0.181	0.018	0.295
converging	0.044	0.001	0.838	0.016	0.189	0.022	0.279
chain	0.022	-0.192	0.248	-0.068	0.095	-0.082	0.110

Table 2: Range of motif frequencies in network samples. “E only” ( or “EI”) means excitatory only networks (or excitatory-inhibitory networks). “SONET” and “degree” refer to the two methods of generating networks.

## D Compensating for fluctuations in empirical connection probability

In order to isolate the impact of higher order motifs, all network samples should have the same empirical connection probability  $p$ , *i.e.* the same number of connections, as we now explain. The algorithms used to generate networks produced samples with slightly different empirical connection probabilities. These fluctuations impact the linear relationship between motif frequencies and average correlation, as  $p$  factors into the regression coefficients (See Eq. (31)). When attempting to determine the regression coefficients from data, fluctuations in  $p$  affect the linear trend.

To address this issue, we can scale motif frequencies by weighting them so that the contribution of a motif to mean correlation does not have a regression coefficient depending on  $p$ . For example, suppose that the theory predicts a linear relation of the form

$$\frac{\langle \tilde{C}^\infty \rangle}{\tilde{C}^0} = f_0(p) + f_{ch}(p)q_{ch} + f_{div}(p)q_{div}.$$

Then we may define auxillary motif frequencies

$$q'_x = \frac{f_x(p)}{f_x(p_{stat})}q_x,$$

and also replace  $f_0(p)$  with  $f_0(p_{stat})$  so that the theoretically-predicted regression relationship becomes

$$\frac{\langle \tilde{C}^\infty \rangle}{\tilde{C}^0} = f_0(p_{stat}) + f_{ch}(p_{stat})q'_{ch} + f_{div}(p_{stat})q'_{div}.$$

Here, we have adjusted the definition of a motif frequency in order to account for finite-size fluctuations in the empirical connection probability  $p$ . The linear regression fit to the quantities  $q'_{\text{ch}}, q'_{\text{div}}$  is much improved, achieving an  $R^2$  measure of 0.99, up from 0.8 in the case where we did not account for such fluctuations. In Fig. 15, we show the scatter plots exploring the relationship between motifs and mean correlation after performing this scaling, and observe the same key trends of mean correlation as in the unadjusted Fig 5: strong and positive dependence on chain and diverging motifs, and minimal dependence on converging motif. These trends are now presented even more clearly.

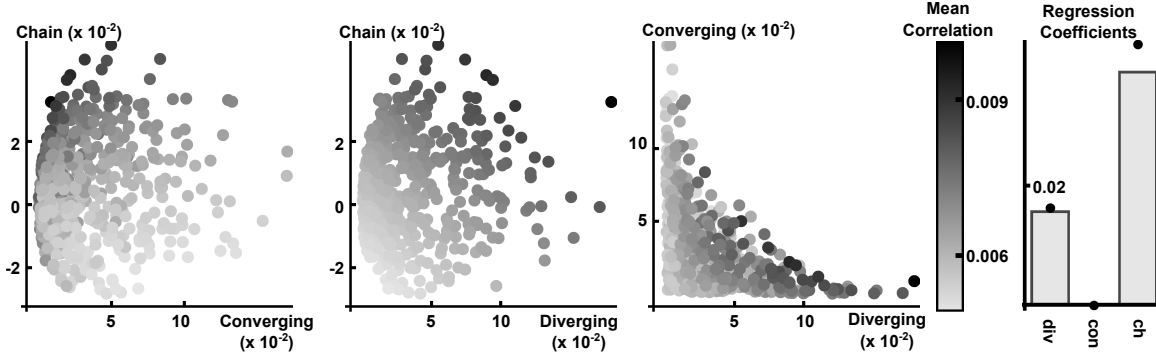


Figure 15: Reproduction of Fig. 5 where we have scaled motifs to account for fluctuations in  $p$ , as described in the text.

## E Proof of Proposition 4.1

We will make use of the following lemma, which may be verified by direct computation.

**Lemma E.1.** *Let  $\{x_n\}_{n \geq 1}$ ,  $\{y_m\}_{m \geq 1}$ ,  $\{z_{nm}\}_{n,m \geq 1}$  be sequences which converge absolutely when summed, and also satisfy*

$$\left| \sum_{n=1}^{\infty} x_n \right| < 1, \quad \left| \sum_{m=1}^{\infty} y_m \right| < 1.$$

*Then,*

$$\begin{aligned} \sum_{i=1}^{\infty} \sum_{(n_1, \dots, n_k) \in \{i\}} \left( \prod_{s=1}^k x_{n_s} \right) &= \sum_{i=1}^{\infty} \left( \sum_{n=1}^{\infty} x_n \right)^i, \\ \sum_{i,j=1}^{\infty} \sum_{\substack{(n_1, \dots, n_{k+1}) \in \{i\} \\ (m_1, \dots, m_{l+1}) \in \{j\}}} \left[ \left( \prod_{s=1}^k x_{n_s} \right) z_{n_{k+1}m_{l+1}} \left( \prod_{t=1}^l y_{m_t} \right) \right] \\ &= \left[ \sum_{i=0}^{\infty} \left( \sum_{n=1}^{\infty} x_n \right)^i \right] \left( \sum_{n,m=1}^{\infty} z_{nm} \right) \left[ \sum_{j=0}^{\infty} \left( \sum_{m=1}^{\infty} y_m \right)^j \right]. \end{aligned}$$

where the sum over  $(n_1, \dots, n_k) \in \{i\}$  denotes a sum over all ordered partitions of  $i$ ,  $(n_1, \dots, n_k)$ , of length  $1 \leq k \leq i$ , with each  $n_l \geq 1$ .

A general result is that, for any matrix  $\mathbf{A}$ , there exists a sub-multiplicative matrix norm ( $\|\mathbf{XY}\| \leq \|\mathbf{X}\| \|\mathbf{Y}\|$ ) arbitrarily approaching the spectral radius  $\Psi(\mathbf{A})$ , that is  $\|\mathbf{A}\| < \Psi(\mathbf{A}) + \epsilon$  [Horn and Johnson, 1990]. Since  $\Psi(\mathbf{K}\Theta) < 1$ , we can choose a sub-multiplicative matrix norm  $\|\cdot\|_\lambda$  satisfying  $\|\mathbf{K}\Theta\|_\lambda < 1$ . As all matrix norms are equivalent, there also exists a constant  $c$  satisfying  $\|\cdot\|_2 \leq c\|\cdot\|_\lambda$  [Horn and Johnson, 1990]. For  $n \geq 1$ , we have

$$\begin{aligned} |\mathbf{u}^T \mathbf{K}_n \mathbf{u}| &= |\mathbf{u}^T (\mathbf{K}\Theta)^{n-1} \mathbf{K} \mathbf{u}| \\ &\leq \|\mathbf{u}\|_2 \cdot \|(\mathbf{K}\Theta)^{n-1}\|_2 \cdot \|\mathbf{K}\|_2 \|\mathbf{u}\|_2 \\ &\leq c \|(\mathbf{K}\Theta)^{n-1}\|_\lambda \|\mathbf{K}\|_2 \\ &\leq c \|\mathbf{K}\|_2 \|\mathbf{K}\Theta\|_\lambda^{n-1}. \end{aligned}$$

Thus  $\sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u}$  converges absolutely, as it is bounded above in absolute value by a convergent geometric series. Since  $\mathbf{u}^T \mathbf{K}_n \mathbf{u} = \mathbf{u}^T \mathbf{K}_n^T \mathbf{u}$ , this also implies the convergence of  $\sum_{m=1}^{\infty} \mathbf{u}^T \mathbf{K}_m^T \mathbf{u}$ . Absolute convergence of  $\sum_{n,m=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \Theta \mathbf{K}_m^T \mathbf{u}$  is proved similarly by noting that for  $n, m \geq 1$ ,

$$\begin{aligned} |\mathbf{u}^T \mathbf{K}_n \Theta \mathbf{K}_m^T \mathbf{u}| &= |\mathbf{u}^T (\mathbf{K}\Theta)^n \mathbf{K}^T (\Theta \mathbf{K}^T)^{m-1} \mathbf{u}| \\ &\leq \|\mathbf{u}\|_2 \|(\mathbf{K}\Theta)^n\|_2 \|\mathbf{K}^T\|_2 \|(\Theta \mathbf{K}^T)^{m-1}\|_2 \|\mathbf{u}\|_2 \\ &\leq c^2 \|\mathbf{K}\|_2 \|(\mathbf{K}\Theta)^n\|_\lambda \|(\Theta \mathbf{K}^T)^{m-1}\|_\lambda \\ &\leq c^2 \|\mathbf{K}\|_2 \|\mathbf{K}\Theta\|_\lambda^{n+m-1}. \end{aligned}$$

Now, suppose that  $|\sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u}| < 1$ . We will remove this assumption shortly. We can now expand the right hand side of Eq. (28) as

$$\left[ \sum_{i=0}^{\infty} \left( \sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u} \right)^i \right] \left( 1 + \sum_{i,j \geq 1} \mathbf{u}^T \mathbf{K}_i \Theta \mathbf{K}_j^T \mathbf{u} \right) \left[ \sum_{j=0}^{\infty} \left( \sum_{m=1}^{\infty} \mathbf{u}^T \mathbf{K}_m^T \mathbf{u} \right)^j \right]. \quad (59)$$

Using Lemma E.1, we have that

$$\begin{aligned} (59) &= \left[ \sum_{i=0}^{\infty} \left( \sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u} \right)^i \right] \left[ \sum_{j=0}^{\infty} \left( \sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n^T \mathbf{u} \right)^j \right] \\ &\quad + \left[ \sum_{i=0}^{\infty} \left( \sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u} \right)^i \right] \left( \sum_{i,j \geq 1} \mathbf{u}^T \mathbf{K}_i \Theta \mathbf{K}_j^T \mathbf{u} \right) \left[ \sum_{j=0}^{\infty} \left( \sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n^T \mathbf{u} \right)^j \right] \\ &= \left[ 1 + \sum_{i=1}^{\infty} \sum_{(n_1, \dots, n_k) \in \{i\}} \left( \prod_{s=1}^k \mathbf{u}^T \mathbf{K}_{n_s} \mathbf{u} \right) \right] \left[ 1 + \sum_{j=1}^{\infty} \sum_{(n_1, \dots, n_l) \in \{j\}} \left( \prod_{t=1}^l \mathbf{u}^T \mathbf{K}_{n_t}^T \mathbf{u} \right) \right] \\ &\quad + \sum_{i,j=1}^{\infty} \sum_{\substack{(n_1, \dots, n_{k+1}) \in \{i\} \\ (m_1, \dots, m_{l+1}) \in \{j\}}} \left( \prod_{s=1}^k \mathbf{u}^T \mathbf{K}_{n_s} \mathbf{u} \right) \mathbf{u}^T \mathbf{K}_{n_{k+1}} \Theta \mathbf{K}_{m_{l+1}}^T \mathbf{u} \left( \prod_{t=1}^l \mathbf{u}^T \mathbf{K}_{m_t}^T \mathbf{u} \right). \end{aligned} \quad (60)$$

Next, note that the product term  $\prod_{s=1}^k \mathbf{u}^T \mathbf{K}_{n_s} \mathbf{u}$ , where  $(n_1, \dots, n_k) \in \{i\}$ , is acquired by distributing across sums  $\mathbf{H} + \boldsymbol{\Theta}$  in  $\mathbf{u}^T [\mathbf{K}(\mathbf{H} + \boldsymbol{\Theta})]^{i-1} \mathbf{K} \mathbf{u}$  and taking the unique term in this where factors of  $\mathbf{H} = \mathbf{u} \mathbf{u}^T$  divide the  $i$  factors of  $\mathbf{K}$  into  $k$  blocks  $\mathbf{K}_{n_s}$  of size  $n_s$  joined by factors of  $\boldsymbol{\Theta}$ . Therefore, summing over all possible ordered partitions, we have

$$\mathbf{u}^T \mathbf{K}^i \mathbf{u} = \mathbf{u}^T (\mathbf{K}(\mathbf{H} + \boldsymbol{\Theta}))^{i-1} \mathbf{K} \mathbf{u} = \sum_{(n_1, \dots, n_k) \in \{i\}} \left( \prod_{s=1}^k \mathbf{u}^T \mathbf{K}_{n_s} \mathbf{u} \right). \quad (61)$$

Similarly,

$$\begin{aligned} \mathbf{u}^T \mathbf{K}^i \boldsymbol{\Theta} (\mathbf{K}^T)^j \mathbf{u} &= \mathbf{u}^T (\mathbf{K}(\mathbf{H} + \boldsymbol{\Theta}))^{i-1} \mathbf{K} \boldsymbol{\Theta} \mathbf{K}^T ((\mathbf{H} + \boldsymbol{\Theta}) \mathbf{K}^T)^{j-1} \mathbf{u} \\ &= \sum_{\substack{(n_1, \dots, n_{k+1}) \in \{i\} \\ (m_1, \dots, m_{l+1}) \in \{j\}}} \left( \prod_{s=1}^k \mathbf{u}^T \mathbf{K}_{n_s} \mathbf{u} \right) \mathbf{u}^T \mathbf{K}_{n_{k+1}} \boldsymbol{\Theta} \mathbf{K}_{m_{l+1}}^T \mathbf{u} \left( \prod_{t=1}^l \mathbf{u}^T \mathbf{K}_{m_t}^T \mathbf{u} \right). \end{aligned} \quad (62)$$

Using Eqns. (60-62), we have that

$$\begin{aligned} (59) &= \left( \sum_{i=0}^{\infty} \mathbf{u}^T \mathbf{K}^i \mathbf{u} \right) \left( \sum_{j=0}^{\infty} \mathbf{u}^T (\mathbf{K}^T)^j \mathbf{u} \right) + \sum_{i,j \geq 1} \mathbf{u}^T \mathbf{K}^i \boldsymbol{\Theta} (\mathbf{K}^T)^j \mathbf{u} \\ &= \sum_{i,j \geq 0} \mathbf{u}^T \mathbf{K}^i \mathbf{H} (\mathbf{K}^T)^j \mathbf{u} + \sum_{i,j \geq 0} \mathbf{u}^T \mathbf{K}^i \boldsymbol{\Theta} (\mathbf{K}^T)^j \mathbf{u} \\ &= \sum_{i,j \geq 0} \mathbf{u}^T \mathbf{K}^i (\mathbf{K}^T)^j \mathbf{u} \quad \text{using } \mathbf{I} = \mathbf{H} + \boldsymbol{\Theta} \\ &= \mathbf{u}^T (\mathbf{I} - \mathbf{K})^{-1} (\mathbf{I} - \mathbf{K}^T)^{-1} \mathbf{u}, \quad \text{using } \Psi(\mathbf{K}) < 1, \end{aligned}$$

where we have used  $\mathbf{u}^T \boldsymbol{\Theta} (\mathbf{K}^T)^j \mathbf{u} = \mathbf{u}^T \mathbf{K}^i \boldsymbol{\Theta} \mathbf{u} = 0$  on the second line of the equation.

Lastly, we will eliminate the assumption  $|\sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u}| < 1$ , which was used to establish Eq. (28). To see that this can be done, let  $z$  be a complex number, and replace  $\mathbf{K}$  by  $z\mathbf{K}$  in Eq. (28), giving

$$\begin{aligned} &\mathbf{u}^T (\mathbf{I} - z\mathbf{K})^{-1} (\mathbf{I} - z\mathbf{K}^T)^{-1} \mathbf{u} \\ &= \left( 1 - \sum_{n=1}^{\infty} z^n \mathbf{u}^T \mathbf{K}_n \mathbf{u} \right)^{-1} \left( 1 + \sum_{n,m=1}^{\infty} z^{n+m} \mathbf{u}^T \mathbf{K}_n \boldsymbol{\Theta} \mathbf{K}_m^T \mathbf{u} \right) \left( 1 - \sum_{m=1}^{\infty} z^m \mathbf{u}^T \mathbf{K}_m^T \mathbf{u} \right)^{-1}. \end{aligned} \quad (63)$$

For sufficiently small  $0 < \delta < 1$ ,  $|z| < \delta$  we have that

$$\left| \sum_{n=1}^{\infty} z^n \mathbf{u}^T \mathbf{K}_n \mathbf{u} \right| \leq \delta \sum_{n=1}^{\infty} |\mathbf{u}^T \mathbf{K}_n \mathbf{u}| < 1.$$



The absolute convergence of the series  $\sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u}$ ,  $\sum_{n=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \mathbf{u}$  and  $\sum_{n,m=1}^{\infty} \mathbf{u}^T \mathbf{K}_n \boldsymbol{\Theta} \mathbf{K}_m^T \mathbf{u}$  then guarantees that Eq. (63) holds on  $|z| < \delta$ , and furthermore, that the right hand side of (63) is an analytic function of  $z$  on  $|z| < \delta$ .

Finally, note that the left hand side of (63) is an analytic function of  $z$  on  $|z| < 1/\Psi(\mathbf{K})$ , since the matrix inverses may be expanded as power series on this range. Since the two sides are equal and analytic on  $|z| < \delta$ , the right hand side must also be defined and analytic on  $|z| < 1/\Psi(\mathbf{K})$ , and equal the left hand side on this range. Since  $1/\Psi(\mathbf{K}) > 1$ , we may take  $z = 1$  in Eq. (63), recovering Eq. (28).

## F Proof of Proposition 5.1

The proof of Proposition 5.1 is identical to the one dimensional case Prop. 4.1, except that we need an entrywise argument to eliminate the assumption  $\|\sum_{n=1}^{\infty} \mathbf{U}^T \mathbf{K}_n \mathbf{U}\|_2 < 1$  that enables the expansion of the right hand side of Eq. (41). Let  $z$  be a complex number, and replace  $\mathbf{K}$  by  $z\mathbf{K}$  in Eq. (41), giving

$$\begin{aligned} & \mathbf{U}^T (\mathbf{I} - z\mathbf{K})^{-1} (\mathbf{I} - z\mathbf{K}^T)^{-1} \mathbf{U} \\ &= \left( \mathbf{I} - \sum_{n=1}^{\infty} z^n \mathbf{U}^T \mathbf{K}_n \mathbf{U} \right)^{-1} \left( \mathbf{I} + \sum_{n,m=1}^{\infty} z^{n+m} \mathbf{U}^T \mathbf{K}_n \boldsymbol{\Theta} \mathbf{K}_m^T \mathbf{U} \right) \left( \mathbf{I} - \sum_{m=1}^{\infty} z^m \mathbf{U}^T \mathbf{K}_m^T \mathbf{U} \right)^{-1}. \end{aligned} \quad (64)$$

Note that the series  $\sum_{n=1}^{\infty} \mathbf{U}^T \mathbf{K}_n \mathbf{U}$  absolutely converges in 2-norm. This follows from writing

$$\|\mathbf{U}^T \mathbf{K}_n \mathbf{U}\|_2 \leq \|\mathbf{U}^T\|_2 \|(\mathbf{K}\boldsymbol{\Theta})^{n-1}\|_2 \|\mathbf{K}\|_2 \|\mathbf{U}\|_2 \leq c \|\mathbf{U}^T\|_2 \|\mathbf{K}\boldsymbol{\Theta}\|_{\lambda}^{n-1} \|\mathbf{K}\|_2 \|\mathbf{U}\|_2,$$

and use the assumption that  $\Psi(\mathbf{K}\boldsymbol{\Theta}) < 1$  and one associated sub-multiplicative norm  $\|\mathbf{K}\boldsymbol{\Theta}\|_{\lambda} < 1$ .  $c$  is a constant such that  $\|\cdot\|_2 \leq c\|\cdot\|_{\lambda}$ . For sufficiently small  $0 < \delta < 1$ ,  $\forall |z| < \delta$ , we have that

$$\left\| \sum_{n=1}^{\infty} z^n \mathbf{U}^T \mathbf{K}_n \mathbf{U} \right\|_2 \leq \delta \sum_{n=1}^{\infty} \|\mathbf{U}^T \mathbf{K}_n \mathbf{U}\|_2 < 1.$$

With this condition, we can establish identity (64) in the same way as in the one dimensional case Prop. 4.1 for  $|z| < \delta$ . Furthermore, it is similar as above to show all of the matrix series on the right hand side of Eq. (64) are absolute converge in 2-norm and therefore entry-wise absolute converge. Hence, we have that each entry of the right hand side of Eq. (64) is an analytic function of  $z$  on  $|z| < \delta$ .

On the other hand, the left hand side of Eq. (64) is entry-wise analytic in  $z$  (each entry is indeed a rational function of  $z$ ) on  $|z| < 1/\Psi(\mathbf{K})$ . Since the two sides are equal and analytic on  $|z| < \delta$ , the right hand side must also be defined and analytic on  $|z| < 1/\Psi(\mathbf{K})$ . Note that  $1/\Psi(\mathbf{K}) > 1$ , so we may set  $z = 1$ , yielding Eq. (41).

## G Expression of the linear dependence between $\langle \tilde{\mathbf{C}}_{EE} \rangle$ and motifs

For excitatory-inhibitory networks, the (linearized) relationship between motif frequencies and averaged correlations is given by Eq. (44). Here, we evaluate the terms in this expression in a special case, to yield an explicit expression for the linear dependence of correlation on motif frequencies. We will express the block-wise average covariances  $\langle \tilde{\mathbf{C}}_{EE} \rangle$  in terms of the 20 individual second order motifs. Along with the (approximate) definition of correlation in Eq. (52), this gives the linear weights. We note that this is how we obtain the specific linear weights used in Fig. 3 B.

For simplicity, assume the special case that all connection probabilities are identical ( $p_{XY} = p$  for all  $X, Y \in \{E, I\}$ ). Define the weight of all excitatory (resp. inhibitory) connections into a cell as

$$\mu_E = pN_E w_E \quad (\text{resp. } \mu_I = pN_I w_I)$$

and the new weight of all connections into a cell as

$$\mu = \mu_E + \mu_I.$$

In addition, define  $\eta$  to be the strength of total common input to a cell pair in an Erdős-Rényi network:

$$\eta = N_E w_E^2 p^2 + N_I w_I^2 p^2.$$

Noting that we have

$$(\mathbf{M}\mathbf{D}_2)^2 = \mu (\mathbf{M}\mathbf{D}_2),$$

it is simple to show that

$$\left( \mathbf{I} - \tilde{A}\mathbf{M}\mathbf{D}_2 \right)^{-1} = \mathbf{I} + \frac{\tilde{A}}{1 - \tilde{A}\mu} (\mathbf{M}\mathbf{D}_2).$$

Similarly

$$\left( \mathbf{I} - \tilde{A}\mathbf{D}_2\mathbf{M} \right)^{-1} = \mathbf{I} + \frac{\tilde{A}}{1 - \tilde{A}\mu} (\mathbf{D}_2\mathbf{M}).$$

Then, from Eq. (44) we get the linear dependence on motifs (treat  $p$  as fixed) as

$$\begin{aligned}
\langle \tilde{\mathbf{C}}_{EE} \rangle / \tilde{C}^0 \sim & \tilde{A}^2 \left( \frac{1 - \tilde{A}\mu_I}{1 - \tilde{A}\mu} \right)^2 Q_{\text{div}}^{EE} + \tilde{A}^2 \frac{\tilde{A}\mu_I(1 - \tilde{A}\mu_I)}{(1 - \tilde{A}\mu)^2} Q_{\text{div}}^{EI} + \tilde{A}^2 \left( \frac{\tilde{A}\mu_I}{1 - \tilde{A}\mu} \right)^2 Q_{\text{div}}^{II} \\
& + 2\tilde{A}^2 \frac{1 - \tilde{A}\mu_I}{1 - \tilde{A}\mu} \frac{(1 - \tilde{A}\mu_I)^2 + \tilde{A}^2 \mu_I^2 \frac{N_E}{N_I}}{(1 - \tilde{A}\mu)^2} Q_{\text{ch}}^{EE} \\
& + 2\tilde{A}^2 \frac{1 - \tilde{A}\mu_I}{1 - \tilde{A}\mu} \frac{(1 - \tilde{A}\mu_E)\tilde{A}\mu_I + (1 - \tilde{A}\mu_I)\tilde{A}\mu_E \frac{N_I}{N_E}}{(1 - \tilde{A}\mu)^2} Q_{\text{ch}}^{EI} \\
& + 2\tilde{A}^2 \frac{\tilde{A}\mu_I}{1 - \tilde{A}\mu} \frac{(1 - \tilde{A}\mu_I)^2 + \tilde{A}^2 \mu_I^2 \frac{N_E}{N_I}}{(1 - \tilde{A}\mu)^2} Q_{\text{ch}}^{IE} \\
& + 2\tilde{A}^2 \frac{\tilde{A}\mu_I}{1 - \tilde{A}\mu} \frac{(1 - \tilde{A}\mu_E)\tilde{A}\mu_I + (1 - \tilde{A}\mu_I)\tilde{A}\mu_E \frac{N_I}{N_E}}{(1 - \tilde{A}\mu)^2} Q_{\text{ch}}^{II},
\end{aligned} \tag{65}$$

where  $Q_{\text{div}}^{XY}, Q_{\text{ch}}^{XY}$  were defined in Eqns. (36,38) as

$$Q_{\text{div}}^{XY} = N_E w_E^2 q_{\text{div}}^{XY,E} + N_I w_I^2 q_{\text{div}}^{XY,I}, \quad Q_{\text{ch}}^{XY} = N_E w_X w_E q_{\text{ch}}^{XEY} + N_I w_X w_I q_{\text{ch}}^{XIY}.$$

Since each of the quantities  $Q_{\text{div}}^{XY}, Q_{\text{ch}}^{XY}$  are clearly linear in second order motif frequencies, Eq. (65) gives a linear relation between second order motif frequencies and block-wise averaged correlation in the two population network. Eq. (65) is the two population analog of Eq. (31) in the single population case.

## H Intuition for why the resumming approach can produce accurate results

From Eq. (29) and Eq. (42), we see that the error of resumming theory is determined by the tail that we dropped in two series. With respect to the coupling strength order of magnitude ( $w$  or  $w_E, w_I$ ), these are geometric series. Therefore the sum of the tail series is controlled by the leading term, that is

$$\tilde{A}^3 \mathbf{L}^T \mathbf{W} \Theta \mathbf{W} \Theta \mathbf{W} \mathbf{L}, \quad \tilde{A}^3 \mathbf{L}^T \mathbf{W} \Theta \mathbf{W}^T \Theta \mathbf{W} \mathbf{L}, \quad \tilde{A}^3 \mathbf{L}^T \mathbf{W} \Theta \mathbf{W} \Theta \mathbf{W}^T \mathbf{L},$$

This shows that the resumming theory has third order accuracy in the effective interaction strength  $\tilde{A}w$ .

Another important factor affecting the accuracy of the resumming theory is the spectral radius  $\Psi(\Theta \mathbf{W} \Theta) = \Psi(\mathbf{W} \Theta) = \Psi(\Theta \mathbf{W})$  (equalities follow from writing  $\mathbf{W}$  under basis of projections  $\Theta$  and  $\mathbf{H}$ ), which is related to how fast the terms in the series converge to 0. There is a simple intuition of  $\Psi(\mathbf{W} \Theta)$  for Erdős-Rényi networks. For single population networks, note that (asymptotically, for large  $N$ )  $\mathbf{W}^0 \Theta = \mathbf{W}^0 - \mathbf{W}^0 \mathbf{H}$  is a matrix with i.i.d. entries of mean 0. According to the Circular Law, all eigenvalues will asymptotically

uniformly distributed within a circle about the origin. Comparing with the spectra of  $\mathbf{W}^0$  in Sec. 2.4, multiplying by  $\Theta$  effectively removes the single dominant eigenvalue [Rajan and Abbott, 2006].

The removal of this dominant eigenvalue will reduce the spectral radius of  $\mathbf{W}^0\Theta$  as compared to  $\mathbf{W}^0$  by a factor of  $\sqrt{N}$  ( $w\sqrt{p(1-p)N}$  compared to  $pNw$ , see Sec. 2.4). Such reductions of  $\Psi(\mathbf{W}\Theta)$  approximately occur in single population networks and at the blocks of  $\mathbf{W}\Theta$  in two population networks, even though those networks are non-Erdős-Rényi. This intuition may help understand why resumming theory works much better than truncation theory: the tails of series in  $\mathbf{W}\Theta$  may be much lighter than those in  $\mathbf{W}$ .

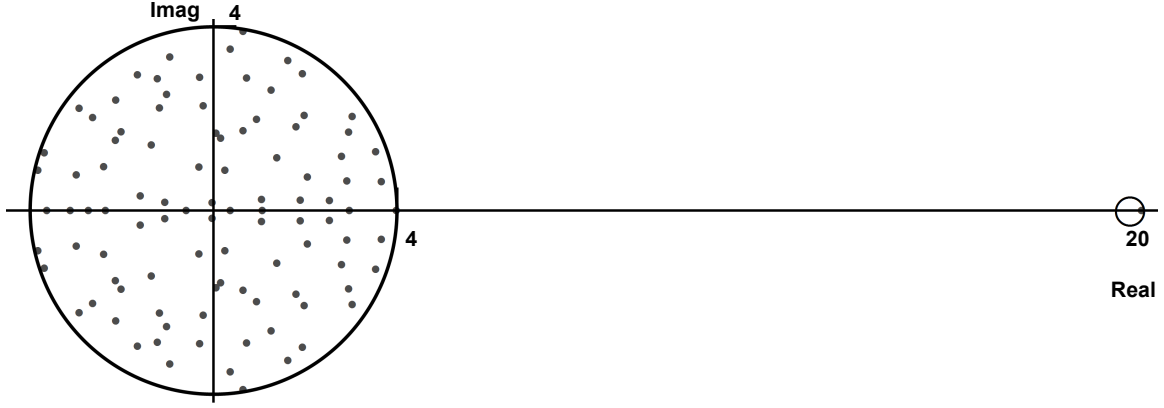


Figure 16: The spectra of single population Erdős-Rényi network, the larger circle and small circle on the right are expected locations of the bulk spectra and single eigenvalue calculated from the asymptotic formula in text above and Sec. 2.4. The excitatory only network has 100 neurons, and  $p = 0.2$ ,  $w = 1$ .

Symbol	Description
$v_i, \tau_i, E_{L,i}, \sigma_i$	Membrane potential, membrane time constant, leak reversal potential, and noise intensity of cell $i$ .
$E_i, \sigma_i$	Mean and standard deviation of the background noise for cell $i$ .
$v_{th}, v_r, \tau_{ref}$	Membrane potential threshold, reset, and absolute refractory period for cells.
$\psi(v), v_T, \Delta_T$	Spike generating current, soft threshold and spike shape parameters for the IF model Fourcaud-Trocmé et al. [2003].
$f_i(t), \eta_i(t)$	Synaptic input from other cells in the network, and external input to cell $i$ .
$\tau_{S,i}, \tau_{D,i}$	Synaptic time constant and delay for outputs of cell $i$ .
$y_i(t)$	Spike train of cell $i$ .
$\mathbf{W}_{ij}$	The $j \rightarrow i$ synaptic weight, proportional to the area under a single post-synaptic current for current-based synapses.
$\mathbf{J}_{ij}(t)$	The $j \rightarrow i$ synaptic kernel - equals the product of the synaptic weight $\mathbf{W}_{ij}$ and the synaptic filter for outputs of cell $j$ .
$\mathbf{C}_{ij}(\tau)$	The cross-correlation function between cells $i, j$ defined by $C_{ij}(\tau) = \text{cov}(y_i(t + \tau), y_j(t))$ .
$\tilde{\mathbf{C}} = \tilde{\mathbf{C}}(0)$	The cross-spectrum matrix evaluated at 0 frequency. Unless noted otherwise, all spectral quantities are evaluated at 0 frequency.
$N_{y_i}(t, t + \tau), \rho_{ij}(\tau)$	Spike count for cell $i$ , and spike count correlation coefficient for cells $i, j$ over windows of length $\tau$ .
$r_i, A_i(t), \mathbf{C}_{ii}^0$	Stationary rate, linear response kernel and uncoupled auto-correlation function for cell $i$ .
$\mathbf{K}_{ij}(t)$	The $j \rightarrow i$ interaction kernel - describes how the firing activity of cell $i$ is perturbed by an input spike from cell $j$ . It is defined by $\mathbf{K}_{ij}(t) = (A_i * \mathbf{J}_{ij})(t)$ .
$\mathbf{y}_i^n(t), \mathbf{C}_{ij}^n(t)$	The $n^{th}$ order approximation of the activity of cell $i$ in a network which accounts for directed paths through the network graph up to length $n$ ending at cell $i$ , and the cross-correlation between the $n^{th}$ order approximations of the activity of cells $i, j$ .
$g(t), \tilde{g}(\omega)$	$\tilde{g}(\omega)$ is the Fourier transform of $g(t)$ with the convention $\tilde{g}(\omega) = \mathcal{F}[g](\omega) \equiv \int_{-\infty}^{\infty} e^{-2\pi i \omega t} g(t) dt$
$\mathbf{E}_e[\cdot]$	Empirical average, $\frac{1}{N^2} \sum_{i,j} \cdot$ or $\frac{1}{N^3} \sum_{i,j,k} \cdot$ depending on context
$R^2$	coefficient of determination, i.e. the square of correlation coefficient $\rho^2$
$\mathbf{W}^0$	Adjacency matrix
$\mathbf{H}$	$\mathbf{H} = \frac{1}{N} \mathbf{1}_{NN}$
$\Theta$	$\Theta = \mathbf{I} - \mathbf{H}$

Table 3: Notation used in the text.

## Acknowledgements

We thank Chris Hoffman and Brent Doiron for their helpful insights. This work was supported by NSF grants DMS-0817649, DMS-1122094, and a Texas ARP/ATP award to KJ, and by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and NSF Grants DMS-1056125 and DMS-0818153 to ESB.

## References

- L F Abbott and Peter Dayan. The Effect of Correlated Variability on the Accuracy of a Population Code. *Neural. Comput.*, pages 1–11, March 1999.
- W. Bair, E. Zohary, and W.T. Newsome. Correlated firing in macaque visual area mt: time scales and relationship to behavior. *J Neurosci*, 21(5):1676–1697, 2001.
- J. Beck, V.R. Bejjanki, and A. Pouget. Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. *Neural Comput*, 23(6):1484–1502, 2011.
- N. Brunel, F.S. Chance, N. Fourcaud, and LF Abbott. Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys Rev Lett*, 86(10):2186–2189, 2001.
- Randy M Bruno. Synchrony in sensation. *Curr. Opin. Neurobiol.*, 21(5):701–8, Oct 2011.
- F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Ann. of Comb.*, 6(2):125–145, 2002.
- M.R. Cohen and A. Kohn. Measuring and interpreting neuronal correlations. *Nat Neurosci*, 14(7):811–819, 2011.
- Peter Dayan and Larry F. Abbot. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 1st edition, 2001.
- J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448:802–806, 2007.
- B. Doiron, B. Lindner, A. Longtin, L. Maler, and J. Bastian. Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Phys Rev Lett*, 96:048101, 2004.
- A.S. Ecker, P. Berens, G.A. Keliris, M. Bethge, N.K. Logothetis, and A.S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587, 2010.
- A.S. Ecker, P. Berens, A. S. Tolias, and M. Bethge. The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.*, 31:14272–14283, 2011.

- N. Fourcaud-Trocmé, D. Hansel, C. Van Vreeswijk, and N. Brunel. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *J Neurosci*, 23(37):11628–11640, 2003.
- Pascal Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci*, 9(10):474–80, 2005.
- F. Gabbiani and S.J. Cox. *Mathematics for Neuroscientists*. Academic Press, London, 2010.
- S. Haeusler, K. Schuch, and W. Maass. Motif distribution, dynamical properties, and computational performance of two data-based cortical microcircuit templates. *J Physiol Paris.*, 103(1-2):73–87, Jan-Mar 2009.
- A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971a.
- A.G. Hawkes. Point spectra of some mutually exciting point processes. *J Roy Stat Soc B Met*, 33(3):438–443, 1971b.
- R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990.
- K. Josić, E. Shea-Brown, B. Doiron, and J. de La Rocha. Stimulus-dependent correlations and population codes. *Neural Comput*, 21(10):2774–2804, 2009.
- Peter E. Latham and Sheila Nirenberg. Synergy, Redundancy, and Independence in Population Codes, Revisited. *J Neurosci*, 25(21):5195–5206, 2005.
- B. Lindner and L. Schimansky-Geier. Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys Rev Lett*, 86(14):2934–2937, 2001.
- B. Lindner, B. Doiron, and A. Longtin. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys Rev E*, 72(6):061919, 2005.
- D. Marinazzo, H.J. Kappen, and S.C.A.M. Gielen. Input-driven oscillations in networks with excitatory and inhibitory neurons with dynamic synapses. *Neural Comput*, 19(7):1739–1765, 2007.
- M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45:167, 2003.
- Strogatz S. Newman, M. E. J. and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E*, 64(026118), 2001.
- S. Ostojic, N. Brunel, and V. Hakim. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J Neurosci*, 29(33):10234–10253, 2009.

- Srdjan Ostojic and Nicolas Brunel. From Spiking Neuron Models to Linear-Nonlinear Models. *PLoS Comput. Biol.*, 7(1):e1001056, January 2011.
- S. Panzeri, S. Schultz, A. Treves, and E. Rolls. Correlations and encoding of information in the nervous system. *P Roy Soc Lond B Bio*, 266:1001–1012, 1999.
- S. Panzeri, R.S. Petersen, S.R. Schultz, M. Lebedev, and M.E. Diamond. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29:769–777, 2001.
- V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. Recurrent interactions in spiking networks with arbitrary topology. *Phys Rev E*, 85(031916), Mar 2012.
- Volker Pernice, Benjamin Staude, Stefano Cardanobile, and Stefan Rotter. How structure determines correlations in neuronal networks. *PLoS Comput Biol*, 7(5):e1002059, 05 2011. doi: 10.1371/journal.pcbi.1002059.
- K Rajan and L. F. Abbott. Eigenvalue spectra of random matrices for neural networks. *Phys Rev Lett*, 97:188104, 2006.
- A.V. Rangan. Diagrammatic expansion of pulse-coupled network dynamics. *Phys Rev Lett*, 102(15):158101, 2009a.
- A.V. Rangan. Diagrammatic expansion of pulse-coupled network dynamics in terms of subnetworks. *Phys Rev E*, 80(3):036101, 2009b.
- A Renart, N Brunel, and XJ Wang. Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In J Feng, editor, *Computational Neuroscience: A Comprehensive Approach*, pages 431–490. CRC Press, Boca Raton, 2004.
- A. Renart, J. de la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K.D. Harris. The asynchronous state in cortical circuits. *Science*, 327(5965):587–590, 2010.
- M.J.E. Richardson. Dynamics of populations and networks of neurons with voltage-activated and calcium-activated currents. *Phys Rev E*, 80(2):021928, 2009.
- H. Risken. *The Fokker-Planck equation: Methods of solution and applications*. Springer Verlag, Berlin, 1996.
- R. Rosenbaum and K. Josić. Mechanisms that modulate the transfer of spiking correlations. *Neural Comput*, 23(5):1261–1305, 2011.
- A. Roxin. The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. *Front Comput Neurosci*, 5(8):1–15, 2011.
- E. Salinas and T. J. Sejnowski. Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J Neurosci*, 20:6193–6209, 2000.



- Elad Schneidman, William Bialek, and Michael Berry. Synergy, Redundancy, and Independence in Population Codes. *J Neurosci*, 23(37):11539–11553, 2003.
- Elad Schneidman, Micheal J. Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(20):1007–1012, 2006.
- M. N. Shadlen and W. T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.*, 18:3870–3896, 1998a.
- M.N. Shadlen and W.T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci*, 18(10):3870–3896, 1998b.
- E. Shea-Brown, K. Josić, J. de La Rocha, and B. Doiron. Correlation and synchrony transfer in integrate-and-fire neurons: Basic properties and consequences for coding. *Phys Rev Lett*, 100(10):108102, 2008. ISSN 1079-7114.
- J. Shlens, G.D. Field GD, J.L. Gauthier, M.I. Grivich MI, D. Petrusca D, A. Sher A, A.M. Litke, and E.J. Chichilnisky EJ. The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.*, 26:8254–8266, 2006.
- W Singer and C M Gray. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.*, 18(1):555–586, 1995.
- H Sompolinsky, H Yoon, K Kang, and M” Shamir. Population coding in neuronal systems with correlated noise. *Phys Rev E*, 64(5 Pt 1):051904, 2001.
- S. Song, P.J. Sjöström, M. Reigl, S. Nelson, and D.B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol*, 3(3):e68, 2005.
- Olaf Sporns and Rolf Kötter. Motifs in brain networks. *PloS Biol*, 2(11):e369, Nov 2004. doi: 10.1371/journal.pbio.0020369.
- R.L. Stratonovich. *Topics in the Theory of Random Noise*, volume 2. Gordon and Breach, New York, 1967.
- T. Tetzlaff, S. Rotter, E. Stark, M. Abeles, A. Aertsen, and M. Diesmann. Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. *Neural Comp*, 20(9):2133–2184, 2008.
- Philipp Khuc Trong and Fred Rieke. Origin of correlated activity between parasol retinal ganglion cells. *Nat Neurosci*, 11(11):1343–1351, September 2008.
- James Trousdale, Yu Hu, Eric Shea-Brown, and Krešimir Josić. Impact of network structure and cellular response on spike time correlations. *PLoS Comput Biol*, 8(3):e1002408, 2012.

- D J Watts and S H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684):440–442, June 1998.
- J A White, J T Rubinstein, and A R Kay. Channel noise in neurons. *Trends Neurosci*, 23 (3):131–137, Mar 2000.
- Liqiong Zhao, Bryce Beverlin, Theoden Netoff, and Duane Q Nykamp. Synchronization from second order network connectivity statistics. *Front Comput Neurosci*, 5:1–16, Jan 2011. doi: 10.3389/fncom.2011.00028.
- E. Zohary, M. N. Shadlen, and W. T. Newsome. Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature*, 370:140–143, 1994.